

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)

Your wildcard search against 2000 terms has yielded the results below

Search for additional matches among the next 2000 terms

starting with: PARSS\$(PARSONS-WINGERTER).P28-P86,P88-P88,P23-P27,P20-P22,P1-P18,P19-P19.

Search Results -

Terms	Documents
128 and pars\$	44

Database:

US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Refine Search:

128 and pars\$

[Clear](#)**Search History****Today's Date: 4/4/2001**

<u>DB Name</u>	<u>Query</u>	<u>Hit Count</u>	<u>Set Name</u>
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l28 and pars\$	44	L29
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l22 and meta same information	95	L28
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l22 and meta information	15	L27
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l21 and multiple party or third party	7699	L26
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l24 and pars\$	15	L25
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l23 and meta same information	17	L24
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l21 and (multiple same party or third same party)	148	L23
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l21 and multiple same party or third same party	8853	L22
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	search engines	1391	L21
USPT	5764906.pn.	1	L20
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l18 and pars\$	20	L19
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l14 and meta same search same engine or third same party same search same engine	30	L18
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	((707/10)!.CCLS.)	1255	L17
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	((707/1)!.CCLS.)	1003	L16
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	((707/5)!.CCLS.)	558	L15
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	((707/3)!.CCLS.)	1173	L14
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	meta search engine or third party search engines	7	L13
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l9 and (whitespace or blank space)	4	L12
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l9 and whitespace or blank space	4007	L11
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l9 and whitespace	0	L10
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l8 and tags	35	L9
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l6 and (terms or words)	56	L8
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l6 and terms or words	547361	L7
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l5 and match\$	57	L6
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l4 and query	63	L5
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l3 and pars\$	87	L4
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and third party or multiple party	476	L3
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	search engines	1391	L2
USPT	search engines	1065	L1

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)**Search Results -**

Terms	Documents
6151584.pn. and parse	1

Database:

US Patents Full-Text Database	▲
US Pre-Grant Publication Full-Text Database	
JPO Abstracts Database	
EPO Abstracts Database	
Derwent World Patents Index	
IBM Technical Disclosure Bulletins	▼

Refine Search:

	▲
	▼

[Clear](#)**Search History****Today's Date: 4/4/2001**

<u>DB Name</u>	<u>Query</u>	<u>Hit Count</u>	<u>Set Name</u>
USPT	6151584.pn. and parse	1	L49
USPT	147 and parse	0	L48
USPT	5826261.pn.	1	L47
USPT	145 and match\$	37	L46
USPT	144 and documents	39	L45
USPT	143 and query	40	L44
USPT	142 and response	50	L43
USPT	141 and pars\$	59	L42
USPT	140 and third party	136	L41
USPT	search engine	1065	L40
USPT	137 and (terms or words)	14	L39
USPT	137 and terms or words	428017	L38
USPT	136 and pars\$	14	L37
USPT	135 and documents	16	L36
USPT	133 and query	16	L35

USPT	l20 and pars\$	0	L34
USPT	third same party same search same engines	20	L33
USPT	l31 and parse	1	L32
USPT	6167393.pn.	1	L31
USPT	l29 and parse	0	L30
USPT	6094649.pn.	1	L29
USPT	l26 and filter\$ same web pages	23	L28
USPT	l25 and (parse or divide)	10	L27
USPT	l25 and parse or divide	131607	L26
USPT	l24 and third same party same search same engines	10	L25
USPT	l22 and quer\$	79	L24
USPT	l22 and query or search	99465	L23
USPT	l21 and meta	85	L22
USPT	search engines	1065	L21
USPT	6094657.pn.	1	L20
USPT	l14 and pars\$	10	L19
USPT	l13 and third same party same search same engines	0	L18
USPT	l14 and multiple same party same search same engines	0	L17
USPT	l14 and third same party same search same engines	0	L16
USPT	l14 and third same party same search same engines or multiple same party search same engines	0	L15
USPT	l13 and query	20	L14
USPT	meta same search same engine	26	L13
USPT	l11 and parse	1	L12
USPT	5913215.pn.	1	L11
USPT	l9 and parse	1	L10
USPT	6078914.pn.	1	L9
USPT	l7 and parse	1	L8
USPT	5933822.pn.	1	L7
USPT	l5 and parse	1	L6
USPT	6151624.pn.	1	L5
USPT	l3 and parse	1	L4
USPT	6012053.pn.	1	L3
USPT	l1 and parse	0	L2
USPT	5926812.pn.	1	L1

File 411:DIALINDEX(R)

DIALINDEX(R)

(c) 2001 The Dialog Corporation plc

*** DIALINDEX search results display in an abbreviated ***
*** format unless you enter the SET DETAIL ON command. ***

?sf compsci,patents,eecomp,electron,allbusiness

You have 329 files in your file list.

(To see banners, use SHOW FILES command)

?show files

File Name

2: INSPEC_1969-2001/APR W1
6: NTIS_1964-2001/Apr W3
8: Ei Compendex(R)_1970-2001/Mar W2
34: SciSearch(R) Cited Ref Sci_1990-2001/Apr W1
35: Dissertation Abstracts Online_1861-2001/Apr
65: Inside Conferences_1993-2001/Apr W1
77: Conference Papers Index_1973-2001/Mar
92: IHS Intl.Stds.& Specs._1999/Nov
94: JICST-EPLUS_1985-2001/MAR W3
99: Wilson Appl. Sci & Tech Abs_1983-2001/Feb
103: Energy SciTec_1974-2001/Mar B2
108: AEROSPACE DATABASE_1962-2001/MAR
144: PASCAL_1973-2001/APR W1
202: INFORMATION SCIENCE ABS._1966-2001/ISSUE 02
233: Internet & Personal Comp. Abs._1981-2001/Apr
238: Abs. in New Tech & Eng._1981-2001/Mar
239: Mathsci_1940-2001/May
275: Gale Group Computer DB(TM)_1983-2001/Apr 04
434: SciSearch(R) Cited Ref Sci_1974-1989/Dec
647: CMP COMPUTER FULLTEXT_1988-2001/APR W1
674: Computer News Fulltext_1989-2001/Mar W3
696: DIALOG Telecom. Newsletters_1995-2001/Apr 04
123: CLAIMS(R)/CURRENT LEGAL STATUS_1980-2001/MAR 27
340: CLAIMS(R)/US PATENT_1950-01/Apr 03
342: Derwent Patents Citation Indx_1978-01/200115
344: CHINESE PATENTS ABS_APR 1985-2001/Feb
345: Inpadoc/Fam.& Legal Stat_1968-2001/UD=200112
347: JAPIO_Oct 1976-2000/Nov(UPDATED 010309)
348: EUROPEAN PATENTS_1978-2001/Mar W04
349: PCT Fulltext_1983-2001/UB=20010329, UT=20010315
371: FRENCH PATENTS_1961-2001/BOPI 200113
447: IMSWorld Patents International_2001/Mar
652: US Patents Fulltext_1971-1979
653: US Patents Fulltext_1980-1989
654: US PAT.FULL._1990-2001/APR 03
670: LitAlert_1973-2001/UD=200113
241: Elec. Power DB_1972-1999Jan
9: Business & Industry(R)_Jul/1994-2001/Apr 04
15: ABI/Inform(R)_1971-2001/Apr 04
16: Gale Group PROMT(R)_1990-2001/Apr 04
18: Gale Group F&S Index(R)_1988-2001/Apr 04
20: World Reporter_1997-2001/Apr 05
148: Gale Group Trade & Industry DB_1976-2001/Apr 04

160: Gale Group PROMT(R)_1972-1989
256: SoftBase:Reviews,Companies&Prods._85-2001/Feb
481: DELPHES EUR BUS_80-1999/DEC W3
583: Gale Group Globalbase(TM)_1986-2001/Apr 05
621: Gale Group New Prod.Annou.(R)_1985-2001/Apr 04
624: MCGRAW-HILL PUBLICATIONS_1985-2001/APR 03
635: Business Dateline(R)_1985-2001/Apr 04
636: Gale Group Newsletter DB(TM)_1987-2001/Apr 04
7: Social SciSearch(R)_1972-2001/Apr W1
13: BAMP_2001/Mar W4
19: CHEM.INDUSTRY NOTES_1974-2001/ISS 200114
22: Employee Benefits_1986-2001/Apr
26: Foundation Directory_2000/Dec
27: Foundation Grants Index_1990-2001/Mar
30: AsiaPacific_1985-2001/Mar 13
33: Aluminium Ind Abs_1968-2001/Apr
42: PHARMACEUTICAL NEWS INDEX_1974-2001/Apr W1
43: Health News Daily_1990-2001/Mar 23
47: Gale Group Magazine DB(TM)_1959-2001/Apr 04
49: PAIS Int._1976-2001/Feb
50: CAB Abstracts_1972-2001/Feb
54: FOODLINE(R): Market Data_1979-2001/Apr 05
63: Transport Res(TRIS)_1970-2001/Feb
67: World Textiles_1968-2001/Mar
73: EMBASE_1974-2001/Mar W4
75: TGG MANAGEMENT CONTENTS(R)_86-2001/MAR W4
79: Foods Adlibra(TM)_1974-2001/Mar
80: TGG Aerospace/Def.Mkts(R)_1986-2001/Apr 04
85: GRANTS_2001/APR
93: TableBase(R) Sep_1997-2001/Mar W4
111: TGG Natl.Newspaper Index(SM)_1979-2001/Apr 02
112: UBM Industry News_1998-2001/Apr 05
113: European R&D Database_1997
115: Research Centers & Services_1994-2000/Nov
116: Brands & Their Companies_2000/Dec
119: Textile Technol.Dig._1978-2001/Mar
122: Harvard Business Review_1971-2001/Mar
126: TRADEMARKSCAN(R)-U.K._2001/Mar B2
127: TRADEMARKSCAN(R)-CANADA_2001/MAR 28
129: PHIND(Archival)_1980-2001/Mar W4
130: PHIND(Daily & Current)_2001/Apr 05
131: Pharmacontacts_2001/Mar
145: (Tacoma) The News Tribune_1992-2001/Apr 04
146: Washington Post Online_1983-2001/Apr 03
147: The Kansas City Star_1995-2001/Apr 04
149: TGG HEALTH&WELLNESS DB(SM)_1976-2001/MAR W4
150: Gale Group Legal Res Index(TM)_1980-2001/Apr 05
151: HealthSTAR_1975-2000/Dec
158: DIOGENES(R)_1976-2001/APR W1
167: Medical Device Register (R)_1999
168: Healthcare Organizations_1999
169: Insurance Periodicals_1984-1999/Nov 15
177: Adv.& Agency Red Books:Advertisers_2001/Mar
178: Adv.& Agency Red Books:Agencies_2001/Mar
180: Federal Register_1985-2001/Apr 04
187: F-D-C Reports_1987-2001/Apr W1
188: Health Devices Sourcebook_(2000)

192: Industry Trends & Anal._1997/Jun
194: CBD_1982/DEC-2000/DEC
195: CBD_Jan 2000-2001/Apr 06
196: FINDEX_1982-1999/Q2
211: Gale Group Newsearch(TM)_2001/Apr 04
226: TRADEMARKSCAN(R)-US FED_OG010403/AP010321
227: TRADEMARKSCAN(R)- Community Tmks_2001/Mar B2
228: TRADEMARKSCAN(R)-Spain_2001/Mar B2
229: Drug Info._2000/Q3
246: TRADEMARKSCAN(R)-U.S. STATE_2001/Apr 03
248: PIRA_1975-2001Apr W4
252: Packaging Sci&Tech_1982-1997/Oct
258: AP News Jul_2000-2001/Apr 05
261: UPI News_1999-2001/Apr 05
262: CBCA Fulltext_1982-2001/Mar
264: DIALOG Defense Newsletters_1989-2001/Apr 04
267: Finance & Banking Newsletters_2001/Apr 04
268: Banking Information Source_1981-2001/Mar W4
269: Materials Bus.(TM)_1985-2001/Apr
278: Microcomputer Software Guide_2001/Feb
285: BioBusiness(R)_1985-1998/Aug W1
286: BIOCOMMERCE ABS.& DIR._1981-2001/MAR B2
304: The Merck Index Online(SM)_/2000S2
318: Chem-Intell Chem Manu Plnts_1999/Jul
319: Chem Bus NewsBase_1984-2001/Apr 05
321: PLASPEC Materials Select DB_1999/Feb
323: RAPRA Rubber & Plastics_1972-2001/Apr
358: Current BioTech Abs_1983-2001/Jun
359: Chemical Economics Handbook_2000/Jul
360: Specialty Chemicals Update Program_2000/Q2
363: Dir. of Chem. Producers-Products_2000/Q3
364: Dir. of Chem. Producers-Companies_2000/Q3
382: Baton Rouge Advocate_1998-2001/Apr 04
387: The Denver Post_1994-2001/Apr 04
388: PEDS: Defense Program Summaries_1999/May
392: Boston Herald_1995-2001/Apr 04
397: Las Vegas Review-Journal_1997-2001/Apr 05
398: CHEMSEARCH(TM)_1957-2001/MAR
427: Fort Worth Star-Telegram_1993-2001/Apr 03
428: Adis Newsletters(Current)_2001/Apr 06
429: Adis Newsletters(Archive)_1982-2001/Mar 08
432: Tampa Tribune_1998-2001/Apr 03
433: Charleston Newspapers_1997-2001/Apr 04
441: ESPICOM Pharm&Med DEVICE NEWS_2001/Mar W2
443: IMSWorld Pharm. Co. Dir._1982-2001/Q1
445: IMSWorld R&D Focus_1991-2001/Apr W1
446: IMSWorld Product Launches_1982-2001/Mar
449: IMSWorld Company Profiles_1992-2001/Feb
450: Publ., Distr.& Wholesalers_2001/Mar
455: DRUG NEWS & PERSPECTIVES_1992-2001/MAR
458: Daily Essentials_2001/Apr 04
459: Daily Essentials (Archival)_1996-2001/Mar W2
461: USP DI(R) Vol. I_1998/Q3
464: USP DICTIONARY (USAN)_1997
465: Incidence & Prevalence_2001/Q1
468: Public Opinion_1940-2001/Mar W4
471: New York Times Fulltext-90 Day_2001/Apr 05

473: FINANCIAL TIMES ABSTRACTS_1998-2001/APR 02
474: New York Times Abs_1969-2001/Apr 04
475: Wall Street Journal Abs_1973-2001/Apr 04
477: Irish Times_1999-2001/Apr 04
483: Newspaper Abstracts Daily_1986-2001/Apr 03
484: Periodical Abstracts Plustext_1986-2001/Apr W1
485: Accounting & Tax Database_1971-2001/Mar W4
486: Press-Telegram_1992- 2001/Apr 04
487: Columbus Ledger-Enquirer_1994-2001/Mar 25
488: Duluth News-Tribune_1995-2001/Apr 04
489: The News-Sentinel_1991-2001/Apr 04
490: Tallahassee Democrat_1993- 2001/Mar 30
491: CanCorp Canadian Financials_2001/Mar W2
492: Arizona Repub/Phoenix Gaz_1986-2001/Apr 04
494: St LouisPost-Dispatch_1988-2001/Apr 04
497: (Ft.Lauderdale)Sun-Sentinel_1988-2001/Apr 05
498: Detroit Free Press_1987-2001/Apr 04
533: Canadian Bus. Directory_2001/Q1
535: Thomas Register Online(R)_2000/Q4
536: (GARY) POST-TRIBUNE_1992-1999/Dec 30
538: Boca Raton News_1994- 1999/Jul 05
539: Macon Telegraph_1994-2001/Apr 02
550: TFSD IPOs_1980-2001/Apr 04
581: Population Demographics_1999/Mar
582: Augusta Chronicle_1996- 2001/Apr 04
584: KOMPASS USA_2001/Mar
585: KOMPASS Middle East/Africa/Mediterr_2000/Jul
586: KOMPASS Latin America_2000/Nov
587: Jane's Defense&Aerospace_2001/Mar W2
600: Early Edition-U.S._2001/Apr 05
601: Early Edition Canada_2001/Apr 05
603: Newspaper Abstracts_1984-1988
604: Gannett News_1998-2001/Apr 04
605: U.S. Newswire_1999-2001/Apr 05
606: Africa News_1999-2001/Apr 05
607: ITAR/TASS News_1999-2001/Apr 05
608: KR/T Bus.News._1992-2001/Apr 05
609: Bridge World Markets_2000-2001/Apr 05
610: Business Wire_1999-2001/Apr 04
612: Japan Economic Newswire(TM)_1984-2001/Apr 04
613: PR Newswire_1999-2001/Apr 05
614: AFP English Wire_1999-2001/Apr 04
616: Canada NewsWire_1999-2001/Mar 09
617: South American Business Info._1999-2001/Apr 05
618: Xinhua News_1999-2001/Apr 05
619: Asia Intelligence Wire_1995-2001/Apr 04
620: EIU:Viewswire_2001/Apr 04
623: Business Week_1985-2001/Apr W1
625: American Banker Publications_1981-2001/Apr 05
627: EIU: COUNTRY ANALYSIS_2001/APR W1
628: Ctry Risk & Forecasts_2001/APR W1
629: EIU:BUS. NEWSLETTERS_2001/APR W1
630: Los Angeles Times_1993-2001/Apr 03
631: Boston Globe_1980-2001/Apr 04
632: Chicago Tribune_1985-2001/Apr 04
633: Phil.Inquirer_1983-2001/Apr 04
634: SAN JOSE MERCURY_ JUN 1985-2001/Mar 30

637: Journal of Commerce_1986-2001/Apr 04
638: Newsday/New York Newsday_1987-2001/Apr 03
639: The Houston Post_1988-1995/Apr 18
640: San Francisco Chronicle_1988-2001/Apr 04
641: Rocky Mountain News_Jun 1989-2001/Apr 01
642: The Charlotte Observer_1988-2001/Apr 04
643: Grand Forks Herald_1995-2001/Apr 03
644: (Boulder) Daily Camera_1995-2000/Nov 14
645: CONTRA COSTA PAPERS_1995- 2001/Apr 02
646: Consumer Reports_1982-2001/Mar
648: TV and Radio Transcripts_1997-2001/Apr W1
649: Gale Group Newswire ASAP(TM)_2001/Apr 02
657: TRADEMARKSCAN(R)-France_2001/Mar B2
658: TRADEMARKSCAN(R)-Benelux_2001/Mar B2
659: TRADEMARKSCAN(R)-Denmark_2001/Mar B2
660: Federal News Service_1991-2001/Mar 08
661: TRADEMARKSCAN(R)-Switzerland_2001/Mar B2
662: TRADEMARKSCAN(R)-Austria_2001/Mar B2
663: TRADEMARKSCAN(R)-Monaco_2001/Mar B2
665: U.S. Newswire_1995-1999/Apr 29
667: ITAR/TASS News_1996-1999/May 26
671: TRADEMARKSCAN(R)-Intl Register_2001/Mar B2
672: TRADEMARKSCAN(R)-Germany_2001/Mar B2
673: TRADEMARKSCAN(R)-Italy_2001/Mar B2
677: TRADEMARKSCAN(R)-Liechtenstein_2001/Mar B2
683: Omaha World-Herald_1998-2000/Dec 01
684: Bradenton Herald_1992-2001/Apr 04
701: St Paul Pioneer Pr Apr_1988-2001/Apr 01
702: Miami Herald_1983-2001/Apr 03
703: USA Today_1989-2001/Apr 04
704: (Portland)The Oregonian_1989-2001/Mar 29
705: The Orlando Sentinel_1988-2001/Apr 05
706: (New Orleans)Times Picayune_1989-2000/Sep 15
707: The Seattle Times_1989-2001/Apr 03
708: Akron Beacon Journal_1989-2001/Apr 03
709: Richmond Times-Disp._1989-2001/Apr 04
710: Times/Sun.Times(London)_Jun 1988-2001/Apr 04
711: Independent(London)_Sep 1988-2001/Apr 04
712: Palm Beach Post_1989-2001/Apr 01
713: Atlanta J/Const._1989-2001/Apr 04
714: (Baltimore) The Sun_1990-2001/Apr 04
715: Christian Sci.Mon._1989-2001/Apr 05
716: Daily News Of L.A._1989-2001/Apr 04
717: The Washington Times_Jun 1989-2001/Apr 03
718: Pittsburgh Post-Gazette_Jun 1990-2001/Apr 05
719: (Albany) The Times Union_Mar 1986-2001/Apr 03
720: (Columbia) The State_Dec 1987-2001/Apr 04
721: Lexington Hrld.-Ldr._1990-2001/Apr 04
722: Cincinnati/Kentucky Post_1990-2001/Mar 31
723: The Wichita Eagle_1990-2001/Apr 04
724: (Minneapolis)Star Tribune_1989-1996/Feb 04
725: (Cleveland)Plain Dealer_Aug 1991-2000/Dec 13
726: S.China Morn.Post_1992--2001/Mar 09
727: Canadian Newspapers_1990-2001/Apr 05
728: ASIA/PAC NEWS_1994-2001/APR W1
731: Philad.Dly.News_1983- 2001/Apr 04
732: San Francisco Exam._1990- 2000/Nov 21

733: The Buffalo News_1990- 2001/Apr 01
734: Dayton Daily News_Oct 1990- 2001/Apr 04
735: St. Petersburg Times_1989- 2000/Nov 01
736: Seattle Post-Int._1990-2001/Apr 03
737: Anchorage Daily News_1989-2001/Apr 04
738: (Allentown) The Morning Call_1990-2001/Apr 04
739: The Fresno Bee_1990-2001/Apr 01
740: (Memphis)Comm.Appeal_1990-2001/Apr 04
741: (Norfolk)Led./Pil._1990-2001/Apr 04
742: (Madison)Cap.Tim/Wi.St.J_1990-2001/Apr 04
743: (New Jersey)The Record_1989-2001/Apr 03
744: (Biloxi) Sun Herald_1995-2001/Mar 04
745: Investext(R) PDF Index_1999--2001/Apr W1
747: Newport News Daily Press_1994-2001/Apr 04
748: Asia/Pac Bus. Jrnls_1994-2001/Apr 04
749: LATIN AMERICAN NEWS JAN/_1994-2001/APR 03
750: Emerging Mkts & Middle East News_1995-2001/Apr 04
753: IBISWORLD MARKET RESEARCH_2000-2001/MAR W4
754: IPO Maven_1994-2000/Jul
755: New Zealand Newspapers_1995-2001/Apr 04
756: Daily/Sunday Telegraph_2000-2001/Apr 04
757: Mirror Publications_2000-2001/Apr 05
758: Asia/Pac Directory_1999/Sep
759: Reuters Business Insight._1992-2001/Mar
760: Euromonitor Strategy_2001/Nov
761: Datamonitor Market Res._1992-2001/Mar
762: Euromonitor Market Res._1991-2001/Feb
763: Freedonia Market Res._1990-2001/Mar
764: BCC Market Research_1989-2001/Mar
765: Frost & Sullivan_1992-1999/Apr
766: (R)Kalorama Info Market Res._1993-2000/Aug
767: Frost & Sullivan Market Eng_2001/Apr
768: EIU Market Research_2001/Jan 26
770: Beverage Marketing Research_2000/Jul
773: EdgarPlus(TM)-Williams Act Filings_2001/Apr 03
774: EdgarPlus(TM)-Prospectuses_2001/Apr 03
775: EdgarPlus(TM)-Reg. Statements_2001/Apr 03
776: EdgarPlus(TM)-6K,8K,& 10C Filings_2001/Apr 03
777: EdgarPlus(TM)-Annual Reports_2001/Apr 03
778: EdgarPlus(TM)-10-K & 20-F Filings_2001/Apr 03
779: EdgarPlus(TM)-10-Q Filings_2001/Apr 03
780: EdgarPlus(TM)-Proxy Statements_2001/Apr 03
781: ProQuest Newsstand_1998-2001/Apr 05
788: (Myrtle Beach) The Sun News_1996-2001/Apr 03
790: Tax Notes Today_1986-2001/Apr 05
791: State Tax Today_1991-2001/Apr 05
792: Worldwide Tax Daily_1987-2001/Apr 05
793: Court Filings_1994-2000/Jan W4
806: Africa News_1996-1999/May 26
810: Business Wire_1986-1999/Feb 28
813: PR Newswire_1987-1999/Apr 30
816: Canada NewsWire_1996-1999/Jun 24
817: South American Business Info._1996-1999/May 24
818: Xinhua News_1996-1999/May 26
861: UPI News_1996-1999/May 27
929: Albuquerque Newspapers_1995-2001/Apr 04
979: Milwaukee Jnl Sentinel_Apr_1998-2001/Apr 03

980: Sarasota Herald-Tribune_1996-2001/Apr 04

?s (multiple(w)search(w)engines) and (kwic or (key(w)word(2w)context))

Your SELECT statement is:

s (multiple(w)search(w)engines) and (kwic or (key(w)word(2w)context))

Items	File
-----	-----
1	349: PCT Fulltext_1983-2001/UB=20010329, UT=20010315
1	654: US PAT.FULL._1990-2001/APR 03
1	15: ABI/Inform(R)_1971-2001/Apr 04
2	148: Gale Group Trade & Industry DB_1976-2001/Apr 04
Examined 50 files	
2	47: Gale Group Magazine DB(TM)_1959-2001/Apr 04
Examined 100 files	
Examined 150 files	
Examined 200 files	
Examined 250 files	
Examined 300 files	

5 files have one or more items; file list includes 329 files.

?begin 349,654

05apr01 10:07:41 User219455 Session D721.2
\$5.96 4.766 DialUnits File411
\$5.96 Estimated cost File411
\$0.30 TYMNET
\$6.26 Estimated cost this search
\$6.50 Estimated total session cost 4.829 DialUnits

SYSTEM:OS - DIALOG OneSearch

File 349:PCT Fulltext 1983-2001/UB=20010329, UT=20010315

(c) 2001 WIPO/MicroPat

File 654:US PAT.FULL. 1990-2001/APR 03

(c) FORMAT ONLY 2001 THE DIALOG CORP.

*File 654: Reassignment data current through 12/5/2000 recordings.

Set	Items	Description
---	-----	-----
?s (multiple(w)search(w)engines) and (kwic or (key(w)word(2w)context))		
	460890	MULTIPLE
	168569	SEARCH
	50190	ENGINES
	33	MULTIPLE(W) SEARCH(W) ENGINES
	32	KWIC
	182495	KEY
	100945	WORD
	128837	CONTEXT
	12	KEY(W) WORD(2W) CONTEXT
S1	2	(MULTIPLE(W) SEARCH(W) ENGINES) AND (KWIC OR (KEY(W) WORD(2W) CONTEXT))

?t 1/2,ab/1-2

1/2,AB/1 (Item 1 from file: 349)
DIALOG(R)File 349:PCT Fulltext

(c) 2001 WIPO/MicroPat. All rts. reserv.

00381374

DATABASE SEARCH SUMMARY WITH USER DETERMINED CHARACTERISTICS
SYNTHESE D'EXPLORATION DE BASES DE DONNEES A CARACTERISTIQUES DETERMINEES
PAR L'UTILISATEUR

Patent Applicant/Assignee:

TELTECH RESOURCE NETWORK CORPORATION

Inventor(s):

THOMSON William K

Patent and Priority Information (Country, Number, Date):

Patent: WO 9512173 A2-A3 19950504

Application: WO 94US11629 19941028 (PCT/WO US9411629)

Priority Application: US 93144767 19931028

Designated States: CA JP AT BE CH DE DK ES FR GB GR IE IT LU MC NL PT SE

Main International Patent Class: G06F-017/30;

Publication Language: English

Fulltext Word Count: 7203

English Abstract

An information storage, searching and retrieval system for large (gigabytes) domains of archived textual data. The system includes multiple query generation processes, a search process, and a presentation of search results that is sorted by category or type and that may be customized based on the professional discipline (or analogous personal characteristic of the user), thereby reducing the amount of time and cost required to retrieve relevant results.

Japanese Abstract

L'invention concerne un systeme de stockage, de recherche et d'extraction d'informations pour de vastes (gigaoctets) domaines de donnees de textes archivees. Ce systeme comprend plusieurs processus de generation d'interrogations, un processus de recherche, et une presentation des resultats de recherches qui sont tries par categorie ou par type. En outre, ces derniers peuvent etre personnalises en fonction de la categorie professionnelle (ou de caracteristiques personnelles analogues de l'utilisateur), ce qui permet de reduire le temps requis et les couts associes a l'extraction des resultats recherches.

1/2,AB/2 (Item 1 from file: 654)

DIALOG(R)File 654:US PAT.FULL.

(c) FORMAT ONLY 2001 THE DIALOG CORP. All rts. reserv.

02652495

Utility

INFORMATION MANAGEMENT SYSTEM

PATENT NO.: 5,634,051

ISSUED: May 27, 1997 (19970527)

INVENTOR(s): Thomson, William K., Spring Valley, OH (Ohio), US (United States of America)

ASSIGNEE(s): Teltech Resource Network Corporation, (A U.S. Company or Corporation), Minneapolis, MN (Minnesota), US (United States of America)

EXTRA INFO: Assignment transaction [Reassigned], recorded February 1, 2000 (20000201)

Assignment transaction [Reassigned], recorded June 30,
2000 (20000630)

APPL. NO.: 8-585,383

FILED: January 11, 1996 (19960111)

This application is a continuation, of application Ser. No. 08-144,767,
filed Oct. 28, 1993, abandoned.

U.S. CLASS: 707-5

INTL CLASS: [6] G06F 17-30

FIELD OF SEARCH: 395-600

References Cited

U.S. PATENT DOCUMENTS

4,542,477	9/1985	Noyori et al.	364-900
4,648,046	3/1987	Copenhaver et al.	395-131
4,703,425	10/1987	Muraki	364-419
4,879,648	11/1989	Cochran et al.	395-275
5,109,509	4/1992	Katayama et al.	395-600
5,151,857	9/1992	Matsui	364-419
5,175,814	12/1992	Anick et al.	395-161
5,197,005	3/1993	Shwartz et al.	364-419
5,241,671	8/1993	Reed et al.	395-600
5,369,778	11/1994	San Soucie et al.	395-800
5,371,673	12/1994	Fan	364-419.1

OTHER REFERENCES

Seaching on Dialog, Dialog Information Services, Inc, Palo Alto, CA, pp.
51, 18, 24, 26, 283, 1992.

Text Search and Retrieval Reference Manual, U.S. Patent and Trademark
Office, Dec. 3, 1991, excerpt pp. p1-p7.

Chen, Hsinchun, et al; "Generating, Integrating, and Activating Thesauri
for Concept-Based Document Retrieval," IEEE Xepert, Apr. 1993, pp. 25-34.

PRIMARY EXAMINER: Amsbury, Wayne

ATTORNEY, AGENT, OR FIRM: Fredrikson & Byron, P.A.

CLAIMS: 22

EXEMPLARY CLAIM: 1

DRAWING PAGES: 5

DRAWING FIGURES: 5

ART UNIT: 237

FULL TEXT: 732 lines

ABSTRACT

An information storage, searching and retrieval system for large
(gigabytes) domains of archived textual dam. The system includes multiple
query generation processes, a search process, and a presentation of search
results that is sorted by category or type and that may be customized based
on the professional discipline (or analogous personal characteristic of the
user), thereby reducing the amount of time and cost required to retrieve
relevant results.

?t 1/2,kwic/1-2

1/2,KWIC/1 (Item 1 from file: 349)
DIALOG(R)File 349:PCT Fulltext
(c) 2001 WIPO/MicroPat. All rts. reserv.

00381374

DATABASE SEARCH SUMMARY WITH USER DETERMINED CHARACTERISTICS
SYNTHESE D'EXPLORATION DE BASES DE DONNEES A CARACTERISTIQUES DETERMINEES
PAR L'UTILISATEUR

Patent Applicant/Assignee:

TELTECH RESOURCE NETWORK CORPORATION

Inventor(s):

THOMSON William K

Patent and Priority Information (Country, Number, Date):

Patent: WO 9512173 A2-A3 19950504

Application: WO 94US11629 19941028 (PCT/WO US9411629)

Priority Application: US 93144767 19931028

Designated States: CA JP AT BE CH DE DK ES FR GB GR IE IT LU MC NL PT SE

Main International Patent Class: G06F-017/30;

Publication Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 7203

Fulltext Availability:

Detailed Description

Detailed Description

... domain and ask if the user wishes to continue. This typically would occur only if multiple search engines are not operational.

If all columns respond or the user indicates that the partial search... category.

Figure 3 refers to a sampling of formats that are possible, such as "short", "KWIC " (key word in context), "abridged" and "complete." Other formats can be utilized as desired. The formats allow the user...

1/2,KWIC/2 (Item 1 from file: 654)
DIALOG(R)File 654:US PAT.FULL.
(c) FORMAT ONLY 2001 THE DIALOG CORP. All rts. reserv.

02652495

Utility

INFORMATION MANAGEMENT SYSTEM

PATENT NO.: 5,634,051

ISSUED: May 27, 1997 (19970527)

INVENTOR(s): Thomson, William K., Spring Valley, OH (Ohio), US (United States of America)

ASSIGNEE(s): Teltech Resource Network Corporation, (A U.S. Company or Corporation), Minneapolis, MN (Minnesota), US (United States of America)

EXTRA INFO: Assignment transaction [Reassigned], recorded February 1,

2000 (20000201)
Assignment transaction [Reassigned], recorded June 30,
2000 (20000630)
APPL. NO.: 8-585,383
FILED: January 11, 1996 (19960111)

This application is a continuation, of application Ser. No. 08-144,767,
filed Oct. 28, 1993, abandoned.

U.S. CLASS: 707-5
INTL CLASS: [6] G06F 17-30
FIELD OF SEARCH: 395-600

References Cited

U.S. PATENT DOCUMENTS

4,542,477	9/1985	Noyori et al.	364-900
4,648,046	3/1987	Copenhaver et al.	395-131
4,703,425	10/1987	Muraki	364-419
4,879,648	11/1989	Cochran et al.	395-275
5,109,509	4/1992	Katayama et al.	395-600
5,151,857	9/1992	Matsui	364-419
5,175,814	12/1992	Anick et al.	395-161
5,197,005	3/1993	Shwartz et al.	364-419
5,241,671	8/1993	Reed et al.	395-600
5,369,778	11/1994	San Soucie et al.	395-800
5,371,673	12/1994	Fan	364-419.1

OTHER REFERENCES

Seaching on Dialog, Dialog Information Services, Inc, Palo Alto, CA, pp.
51, 18, 24, 26, 283, 1992.

Text Search and Retrieval Reference Manual, U.S. Patent and Trademark
Office, Dec. 3, 1991, excerpt pp. p1-p7.

Chen, Hsinchun, et al; "Generating, Integrating, and Activating Thesauri
for Concept-Based Document Retrieval," IEEE Xepert, Apr. 1993, pp. 25-34.

PRIMARY EXAMINER: Amsbury, Wayne
ATTORNEY, AGENT, OR FIRM: Fredrikson & Byron, P.A.
CLAIMS: 22
EXEMPLARY CLAIM: 1
DRAWING PAGES: 5
DRAWING FIGURES: 5
ART UNIT: 237
FULL TEXT: 732 lines

... domain and ask if the user wishes to continue. This typically would
occur only if multiple search engines are not operational.

If all columns respond or the user indicates that the partial search...
category. FIG. 3 refers to a sampling of formats that are possible, such as
"short", " KWIC " (key word in context), "abridged" and "complete."
Other formats can be utilized as desired. The formats allow the user...
?begin 15,148,47

05apr01 10:11:20 User219455 Session D721.3
 \$0.72 0.152 DialUnits File349
 \$2.80 1 Type(s) in Format 2
 \$5.10 1 Type(s) in Format 5 (UDF)
 \$7.90 2 Types
 \$8.62 Estimated cost File349
 \$1.95 0.330 DialUnits File654
 \$0.95 1 Type(s) in Format 2
 \$3.20 1 Type(s) in Format 9 (UDF)
 \$4.15 2 Types
 \$6.10 Estimated cost File654
 OneSearch, 2 files, 0.482 DialUnits FileOS
 \$0.20 TYMNET
 \$14.92 Estimated cost this search
 \$21.42 Estimated total session cost 5.311 DialUnits

SYSTEM:OS - DIALOG OneSearch

File 15:ABI/Inform(R) 1971-2001/Apr 04
 (c) 2001 Bell & Howell
 File 148:Gale Group Trade & Industry DB 1976-2001/Apr 04
 (c)2001 The Gale Group
 File 47:Gale Group Magazine DB(TM) 1959-2001/Apr 04
 (c) 2001 The Gale group

Set	Items	Description
?	s	(multiple(w)search(w)engines) and (kwic or (key(w)word(2w)context))
	524783	MULTIPLE
	379110	SEARCH
	132751	ENGINES
	307	MULTIPLE(W) SEARCH(W) ENGINES
	616	KWIC
	1177917	KEY
	405638	WORD
	191977	CONTEXT
	105	KEY(W)WORD(2W)CONTEXT
S1	5	(MULTIPLE(W)SEARCH(W)ENGINES) AND (KWIC OR (KEY(W)WORD(2W)CONTEXT))
?t	1/2,ab,kwic/1-5	

1/2,AB,KWIC/1 (Item 1 from file: 15)
 DIALOG(R)File 15:ABI/Inform(R)
 (c) 2001 Bell & Howell. All rts. reserv.

01539791 01-90779
 The ASIDIC 1997 fall meeting
 Brenner, Ev
 Information Today v14n10 PP: 15, 58 Nov 1997 ISSN: 8755-6286
 JRNL CODE: IFT
 DOC TYPE: Journal article LANGUAGE: English LENGTH: 2 Pages
 WORD COUNT: 1961

ABSTRACT: The fall meeting of the Association of Information & Dissemination Centers was held in Seattle September 21-23. The theme for the 1 1/2 day meeting, designed and chaired by Harry Collier of Infonortics Ltd. was "Incorporating Intelligent into Networked Information." The

conference provided a look at some controversial topics presented by fine speakers. Sue Lachance spoke of Infoseek's search engine features opening with "Is it the World Wide Web or World Wild Web?" The features she discussed were automatic phrase recognition, proper name recognition, distributed search, topical directories created with neural network NET technology and quality indexing guidelines.

COMPANY NAMES:

Association of Information & Dissemination Centers

GEOGRAPHIC NAMES: US

DESCRIPTORS: Associations; Information retrieval; Meetings; Searches;
Speakers; Online information services; World Wide Web; Electronic
publishing

CLASSIFICATION CODES: 9540 (CN=Nonprofit institutions); 9190 (CN=United
States); 5250 (CN=Telecommunications systems); 8302 (CN=Software and
computer services); 8690 (CN=Publishing industry)

...TEXT: of Infoseek. It has received a patent for a method of searching
the Web via multiple search engines, a technique that is expected to
be fully implemented by the beginning of next year...

...yet dumber at searching when they finally begin to use the computer.

Bellick did a KWIC index of the terms used in the 2,000 queries and
surprisingly found 4,528...

1/2,AB,KWIC/2 (Item 1 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2001 The Gale Group. All rts. reserv.

10156104 SUPPLIER NUMBER: 19952087 (USE FORMAT 7 OR 9 FOR FULL TEXT)
The ASIDIC 1997 fall meeting; speakers focused on search-and-retrieval
technologies and techniques.(Association of Information and Dissemination
Centers)
Brenner, Ev
Information Today, v14, n10, p15(2)
Nov, 1997
ISSN: 8755-6286 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 2080 LINE COUNT: 00167

INDUSTRY CODES/NAMES: BUSN Any type of business; LIB Library and
Information Science

DESCRIPTORS: Association of Information and Dissemination Centers--
Conferences, meetings, seminars, etc.; Information storage and retrieval
systems--Product development

PRODUCT/INDUSTRY NAMES: 7375000 (Database Providers)

SIC CODES: 7375 Information retrieval services

FILE SEGMENT: TI File 148

... of Infoseek. It has received a patent for a method of searching the
Web via multiple search engines, a technique that is expected to be
fully implemented by the beginning of next year...

...yet dumber at searching when they finally begin to use the computer.

Bellick did a KWIC index of the terms used in the 2,000 queries and

surprisingly found 4,528...

1/2,AB,KWIC/3 (Item 2 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2001 The Gale Group. All rts. reserv.

09834434 SUPPLIER NUMBER: 19383933 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Surfing corporate intranets; search tools that control the undertow.
(includes related articles on searching databases via an intranet and
intelligent search agents)
Zorn, Peggy; Emanoil, Mary; Marshall, Lucy; Panek, Mary
Online, v21, n3, p30(16)
May-June, 1997
ISSN: 0146-5422 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 9276 LINE COUNT: 00779

ABSTRACT: The availability of commercial and free intranet search engines promises to meet the unique search and retrieval needs of intranet users. Free and commercial search engine offerings are reviewed based on their technical as well as advanced searching capabilities. Users should choose a search engine based on the type of documents on their site, the site's size, the number of web servers, the server platform and available technical expertise.

SPECIAL FEATURES: table; illustration
INDUSTRY CODES/NAMES: BUSN Any type of business; LIB Library and
Information Science
DESCRIPTORS: Intranets--Usage; Online searching--Usage
PRODUCT/INDUSTRY NAMES: 7399200 (Info Services ex Database)
SIC CODES: 7389 Business services, not elsewhere classified
FILE SEGMENT: TI File 148

... display in native format or just HTML?)
* relevancy ranking or results sorting
* keyword-in-context (KWIC) display
Detailed descriptions of the features and functionality of each product examined follow. They are...lines that matched the query. This results screen thus produces a modified keyword-in-context (KWIC) display, which is extremely useful in determining the relevancy of your retrieval.

Currently, there are...an automatic document summary generator. There is also an option to view the keywords in KWIC mode, where the keywords are highlighted and the user can easily see where the keyword...
<http://www.quarterdeck.com/qdeck/products/webcompass/>) that will not only allow users to query multiple search engines , as is the case with its current release, but will also allow for the inclusion...

1/2,AB,KWIC/4 (Item 1 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2001 The Gale group. All rts. reserv.

05071932 SUPPLIER NUMBER: 19952087 (USE FORMAT 7 OR 9 FOR FULL TEXT)
The ASIDIC 1997 fall meeting; speakers focused on search-and-retrieval technologies and techniques.(Association of Information and Dissemination Centers)

Brenner, Ev
Information Today, v14, n10, p15(2)
Nov, 1997
ISSN: 8755-6286 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 2080 LINE COUNT: 00167

DESCRIPTORS: Association of Information and Dissemination Centers--
Conferences, meetings, seminars, etc.; Information storage and retrieval
systems--Product development
PRODUCT/INDUSTRY NAMES: 7375000 (Database Providers)
SIC CODES: 7375 Information retrieval services
FILE SEGMENT: TI File 148

... of Infoseek. It has received a patent for a method of searching the
Web via multiple search engines, a technique that is expected to be
fully implemented by the beginning of next year...
...yet dumber at searching when they finally begin to use the computer.
Bellick did a KWIC index of the terms used in the 2,000 queries and
surprisingly found 4,528...

1/2,AB,KWIC/5 (Item 2 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2001 The Gale group. All rts. reserv.

04830012 SUPPLIER NUMBER: 19383933 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Surfing corporate intranets; search tools that control the undertow.
(includes related articles on searching databases via an intranet and
intelligent search agents)
Zorn, Peggy; Emanoil, Mary; Marshall, Lucy; Panek, Mary
Online, v21, n3, p30(16)
May-June, 1997
ISSN: 0146-5422 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 9276 LINE COUNT: 00779

ABSTRACT: The availability of commercial and free intranet search engines
promises to meet the unique search and retrieval needs of intranet users.
Free and commercial search engine offerings are reviewed based on their
technical as well as advanced searching capabilities. Users should choose a
search engine based on the type of documents on their site, the site's
size, the number of web servers, the server platform and available
technical expertise.

SPECIAL FEATURES: table; illustration
DESCRIPTORS: Intranets--Usage; Online searching--Usage
PRODUCT/INDUSTRY NAMES: 7399200 (Info Services ex Database)
SIC CODES: 7389 Business services, not elsewhere classified
FILE SEGMENT: TI File 148

... display in native format or just HTML?)
* relevancy ranking or results sorting
* keyword-in-context (KWIC) display
Detailed descriptions of the features and functionality of each
product examined follow. They are...lines that matched the query. This
results screen thus produces a modified keyword-in-context (KWIC)
display, which is extremely useful in determining the relevancy of your
retrieval.

Currently, there are...an automatic document summary generator. There is also an option to view the keywords in KWIC mode, where the keywords are highlighted and the user can easily see where the keyword...
<http://www.quarterdeck.com/qdeck/products/webcompass/>) that will not only allow users to query multiple search engines, as is the case with its current release, but will also allow for the inclusion...
?t 1/9/1-5

1/9/1 (Item 1 from file: 15)
DIALOG(R) File 15:ABI/Inform(R)
(c) 2001 Bell & Howell. All rts. reserv.

01539791 01-90779
The ASIDIC 1997 fall meeting
Brenner, Ev
Information Today v14n10 PP: 15, 58 Nov 1997 ISSN: 8755-6286
JRNL CODE: IFT
DOC TYPE: Journal article LANGUAGE: English LENGTH: 2 Pages
WORD COUNT: 1961

ABSTRACT: The fall meeting of the Association of Information & Dissemination Centers was held in Seattle September 21-23. The theme for the 1 1/2 day meeting, designed and chaired by Harry Collier of Infonortics Ltd. was "Incorporating Intelligent into Networked Information." The conference provided a look at some controversial topics presented by fine speakers. Sue Lachance spoke of Infoseek's search engine features opening with "Is it the World Wide Web or World Wild Web?" The features she discussed were automatic phrase recognition, proper name recognition, distributed search, topical directories created with neural network NET technology and quality indexing guidelines.

TEXT: Headnote:

Speakers focused on search-and-retrieval technologies and techniques

The fall meeting of the Association of Information & Dissemination Centers (ASIDIC) was held in Seattle September 21-23. The theme for the one-and-a-half-day meeting, designed and chaired by Harry Collier of Infonortics, Ltd., was "Incorporating Intelligence into Networked Information." The conference provided a look at some controversial topics presented by fine speakers.

NewsNet Keynoter

What a coup of a keynoter! The printed copy of the speakers' biographies stated, "Andrew Elston is currently evaluating opportunities to continue his career in publishing and information services while he oversees the final closing of operations at NewsNet, Inc. this month."

It appeared that Elston was at ASIDIC to tell us why he thinks NewsNet failed. NewsNet was a 15-year-old, established online database of about 1,000 newsletters and other news formats. It spent a lot of money building interfaces to the Web and went live there just 2 years ago. But when it became clear that NewsNet was no longer a competitive product, the parent company first tried to sell it, then simply gave it up. One important factor was that what appeared on the Web was the same proprietary product as its online version. NewsNet acquired new users on the Net, but the kind that didn't stick around after retrieving a-meaning

one-quick answer. The traditional users who migrated from online to the Web and ordinarily stayed online to obtain an average of 5 documents were now getting an average of 1.2 documents and flitting to other news sites, many of which delivered the same information for free. Since there are many competing sites for news-much of it free-that competition simply began to kill off NewsNet.

Elston rued that NewsNet was the victim of literally not understanding that the behavior of Web users would be very different from that of online users. He noted, in his understandably pessimistic mood, that Knight-Ridder cut off DIALOG, which became Knight-Ridder Information Services, then hung it out to dry, and has since been trying to sell it.

Elston predicted that DIALOG will break up. I also heard this from a respected colleague who predicted the break-up in about 5 years.

Let's Get Some Perspective

But, wait up. I'm not quite sure I go along with these predictions on the basis of the NewsNet experience. I brought my love of historical perspective to bear and thought about what advantages online databases had over printed reference texts back in the 1960s when online was considered revolutionary. Online solved important interdisciplinary problems in the sense that good answers could be obtained from a variety of interdisciplinary reference works-and fast.

Many of the databases had controlled indexing, and skilled professionals could search successfully through Boolean and its sophistications.

That was the value-added aspect lacking in the NewsNet product in its shift to the Web. It had nothing more to offer than its text, so it became just another unindexed site, even though it was probably better organized than most of the free sites. It is understandable that it failed.

But that doesn't mean DIALOG and its databases will fail, fall, or crack up that fast. Databases within DIALOG will fail if they have nothing more to offer than text with no real intelligent added-value features. Will that kill off DIALOG? It depends on whether downsizing in corporate information centers continues unabated. There are indications that, after the downsizing onslaught, there may be a new middle road to the role of intermediary.

Elston spoke of the resurgence of newspapers in print in the 1990s. The recent move of The New York Times in dividing its present version into more sections seems germane. Another interesting point made by Elston concerned venture capital. The venturers are not looking at established companies but to those with uncharted futures. The young, untried technology bucks are looking good. The venturers are now true adventurers.

In a summary wrap-up at the end of a conference on search engines a couple of years ago, one point made by James Callam of the University of Massachusetts was that "Boolean is dead! Long live Boolean!" Recent conferences, including this one, have led me to wonder about "Online is dead! Long live Online!," even though the NewsNet demise was a rather frighteningly speedy one.

IsoQuest

I was intrigued by IsoQuest and its Data Extractopm Technology Tool-Kit. Tony Hall reviewed how far a 15-month-- old software company has come with its NetOwl [see related news announcement in the Internet Publishing Today section on p. 47]. He spoke of an automatic-index software tool that can browse dynamically for company names, place names, and people names, and also can automatically extract pieces of text from full text, thus providing a useful summarization of that article.

Automatic abstracting rears its head once more. Individual, Excalibur, and Infoseek are three retrieval companies that have bought into this product. Entity extraction, the more scholarly name for this kind of retrieval, will be the subject of a major talk presented by the president of IsoQuest, Paul Jacobs, at the "Search Engine and Beyond" conference to be held in Boston April 1-2, 1998.

OCLC's Kilroy Project

Terry Noreault reviewed OCLC's Kilroy Internet database project. OCLC's reaching out to explore the Internet resources at some 800,000 Web sites to establish databases of these resources. Significantly, OCLC is developing statistical means to find these resources and feed them into its traditional Dewey and LC classifications automatically (Scorpion). It is enhancing its classification schema and states that "automatic assignment of classification is feasible." I really don't believe in classification systems for the long run, but apparently, "Classification is dead! Long live Classification!"

Search Engines

Those of you who have been reading Sue Feldman's articles realize what a good communicator she is. Her article in the May 1997 issue of Searcher (p. 44) on search engines is worth reviewing.

In her ASIDIC presentation, Feldman said that searching the Web is good for finding an answer--one good one, that is. If that's what you want, the Web's the place to go. Some may think that's a bit exaggerated, but she wanted to emphasize that that's as far as good searching will go on the Internet. For end users, that's where it's at.

Feldman was particularly good at discussing spamming problems. Other barriers she covered were the size of the Internet, rapid changes of Web pages, and inaccessible text. I liked the fact that she wasn't too enthusiastic about Excite's power-search feature, which was a trend back to Boolean. She stated outright that Boolean is not for Web searching. There are many who naively think that adding Boolean to search engines is a sign of progress. It isn't for end users.

I don't mean to suggest that Feldman is anti-Web. She isn't. She is for improvement and played a devil's advocate's role--much needed at this time.

Infoseek

Sue Lachance spoke of Infoseek's search engine features opening with "Is it the World Wide Web or the World Wild Web?" The features she discussed were automatic phrase recognition, proper name recognition, distributed search, topical directories created with neural network NET technology, and quality indexing guidelines. She had little to say about distributed search, which is an important new development out of Infoseek. It has received a patent for a method of searching the Web via multiple search engines, a

technique that is expected to be fully implemented by the beginning of next year. Infoseek president Steve Kirsch will be addressing this at the Boston meeting in April.

Yes, you've probably guessed it by now. Announcing this Boston meeting is self-serving. I have designed and will chair the program. Please attend anyway. I promise a landmark conference. For more information, contact me or visit the Web site (<http://www.infonortics.com>) and click on "Search Engines Meeting."

The Gorilla Story

Mark Chussil of Advanced Competitive Strategies, Inc. conducts War Games and War Colleges for corporations and other institutions. He is one of my favorite speakers. This was the fourth time I've heard him, and I never tire of his presentations. He was asked to speak because Harry Collier usually designs programs to contain at least one speaker from an allied but remote sphere related to the audience. In this case, we learned a bit about a competitive-intelligence technique.

I paraphrase here his opening story, which teaches us about out-of-the-box thinking, something that's important in these changing times.

In experimenting to see how intelligent a gorilla was, a graduate student shut one up in a room to see how long it would take him to learn to use the doorknob to get out. After a long period of disinterest on the part of the gorilla, the student entered the room to release him, whereupon the gorilla picked up the student and threw him against the wall, thus creating a large hole in the wall through which the gorilla then exited.

We learn three lessons from this story: Never assume that there is only one answer to a question, never assume that you know the best answer, and never assume that you are smarter than a gorilla.

After revealing what simulation is all about and what a corporation goes through in applying itself to War Games and the War College, Chussil ended his talk with an Arnold Palmer quote, "The more I practice, the luckier I get."

If you're at all involved in trying to make major competitive changes and decisions, you should consider Chussil's technique, and you too may get luckier.

Image Retrieval

Gordon Short of Excalibur spoke on advanced techniques in imaging, which are becoming more and more feasible and thus commercial. He spoke of Kanji recognition (recognizing strokes), face recognition (recognizing patterns), scene-- change detection (image similarity/difference), and general-image searching ("gestalt," color-shape-texture).

Excalibur seems quite advanced in the commercialization of these techniques. Short showed a picture of a waterlily, which he used as a reference image, and asked his system to retrieve eight similar objects. The eighth likeness was a bunch of bananas, and one could actually detect the seemingly absurd relationship. However, one could limit the search to flowers and come up with a more relevant set.

Concept-based retrieval of images would be the next revolution of image retrieval and Short predicted it was 3-5 years off. (Brenner's law says to double every prediction you hear or see.)

Smart, Dumb, Dumber

David Bellick is search schema manager of MSN Publishing & Tools of the Microsoft Corporation and has been analyzing 2,000 queries he chose randomly from the Web. He started out by saying that NET IR is still inadequate and is technology driven. We all know that, but what he had to say further was evident, although I hadn't realized it--that the end user on the Net today represents the intelligentsia: the people who are likely to have been to college, the big earners who can afford a computer at home as well as at work.

So now I realize we have three levels of users: 1) the smart professionals who know how to search, 2) the smart end users who are pretty dumb at searching, and 3) a mass of uneducated end users who may possibly be yet dumber at searching when they finally begin to use the computer.

Bellick did a KWIC index of the terms used in the 2,000 queries and surprisingly found 4,528 total terms used and only 2,807 of them unique. The top terms were sex (17.5 percent), computer/Internet (14.9 percent), entertainment (14.1 percent), recreation/leisure (12.7 percent), business/investing (5.6 percent), and medicinal/fitness (4.6 percent).

I wonder how intelligent the intelligentsia really are. Think about it.

Author Affiliation:

Ev Brenner managed the Central Abstracting & Indexing Service of the American Petroleum Institute for 30 years and is now a well-known information industry observer. He can be reached by e-mail at 73632.2644@compuserve.com.

THIS IS THE FULL-TEXT. Copyright Information Today Inc 1997

COMPANY NAMES:

Association of Information & Dissemination Centers

GEOGRAPHIC NAMES: US

DESCRIPTORS: Associations; Information retrieval; Meetings; Searches; Speakers; Online information services; World Wide Web; Electronic publishing

CLASSIFICATION CODES: 9540 (CN=Nonprofit institutions); 9190 (CN=United States); 5250 (CN=Telecommunications systems); 8302 (CN=Software and computer services); 8690 (CN=Publishing industry)

1/9/2 (Item 1 from file: 148)
DIALOG(R) File 148:Gale Group Trade & Industry DB
(c)2001 The Gale Group. All rts. reserv.

10156104 SUPPLIER NUMBER: 19952087 (THIS IS THE FULL TEXT)
The ASIDIC 1997 fall meeting; speakers focused on search-and-retrieval technologies and techniques. (Association of Information and Dissemination Centers)

Brenner, Ev

Information Today, v14, n10, p15(2)

Nov, 1997

ISSN: 8755-6286

LANGUAGE: English

RECORD TYPE: Fulltext

WORD COUNT: 2080

LINE COUNT: 00167

TEXT:

The fall meeting of the Association of Information & Dissemination Centers (ASIDIC) was held in Seattle September 21-23. The theme for the one-and-a-half-day meeting, designed and chaired by Harry Collier of Infonortics, Ltd., was "Incorporating Intelligence into Networked Information." The conference provided a look at some controversial topics presented by fine speakers.

NewsNet Keynoter

What a coup of a keynoter! The printed copy of the speakers' biographies stated, "Andrew Elston is currently evaluating opportunities to continue his career in publishing and information services while he oversees the final closing of operations at NewsNet, Inc. this month."

It appeared that Elston was at ASIDIC to tell us why he thinks NewsNet failed. NewsNet was a 15-year-old, established online database of about 1,000 newsletters and other news formats. It spent a lot of money building interfaces to the Web and went live there just 2 years ago. But when it became clear that NewsNet was no longer a competitive product, the parent company first tried to sell it, then simply gave it up.

One important factor was that what appeared on the Web was the same proprietary product as its online version. NewsNet acquired new users on the Net, but the kind that didn't stick around after retrieving a--meaning one--quick answer. The traditional users who migrated from online to the Web and ordinarily stayed online to obtain an average of 5 documents were now getting an average of 1.2 documents and flitting to other news sites, many of which delivered the same information for free. Since there are many competing sites for news--much of it free--that competition simply began to kill off NewsNet.

Elston rued that NewsNet was the victim of literally not understanding that the behavior of Web users would be very different from that of online users. He noted, in his understandably pessimistic mood, that Knight-Ridder cut off DIALOG, which became Knight-Ridder Information Services, then hung it out to dry, and has since been trying to sell it.

Elston predicted that DIALOG will break up. I also heard this from a respected colleague who predicted the break-up in about 5 years.

Let's Get Some Perspective

But, wait up. I'm not quite sure I go along with these predictions on the basis of the NewsNet experience. I brought my love of historical perspective to bear and thought about what advantages online databases had over printed reference texts back in the 1960s when online was considered revolutionary. Online solved important interdisciplinary problems in the sense that good answers could be obtained from a variety of interdisciplinary reference works--and fast.

Many of the databases had controlled indexing, and skilled professionals could search successfully through Boolean and its sophistications.

That was the value-added aspect lacking in the NewsNet product in its shift to the Web. It had nothing more to offer than its text, so it became just another unindexed site, even though it was probably better organized than most of the free sites. It is understandable that it failed.

But that doesn't mean DIALOG and its databases will fail, fall, or crack up that fast. Databases within DIALOG will fail if they have nothing

more to offer than text with no real intelligent added-value features. Will that kill off DIALOG? It depends on whether downsizing in corporate information centers continues unabated. There are indications that, after the downsizing onslaught, there may be a new middle road to the role of intermediary.

Elston spoke of the resurgence of newspapers in print in the 1990s. The recent move of The New York Times in dividing its present version into more sections seems germane. Another interesting point made by Elston concerned venture capital. The venturers are not looking at established companies but to those with uncharted futures. The young, untried technology bucks are looking good. The venturers are now true adventurers.

In a summary wrap-up at the end of a conference on search engines a couple of years ago, one point made by James Callam of the University of Massachusetts was that "Boolean is dead! Long live Boolean!" Recent conferences, including this one, have led me to wonder about "Online is dead! Long live Online!," even though the NewsNet demise was a rather frighteningly speedy one.

IsoQuest

I was intrigued by IsoQuest and its Data Extraction Technology Tool-Kit. Tony Hall reviewed how far a 15-month-old software company has come with its NetOwl (see related news announcement in the Internet Publishing Today section) He spoke of an automatic-index software tool that can browse dynamically for company names, place names, and people names, and also can automatically extract pieces of text from full text, thus providing a useful summarization of that article.

Automatic abstracting rears its head once more. Individual, Excalibur, and Infoseek are three retrieval companies that have bought into this product. Entity extraction, the more scholarly name for this kind of retrieval, will be the subject of a major talk presented by the president of IsoQuest, Paul Jacobs, at the "Search Engine and Beyond" conference to be held in Boston April 1-2, 1998.

OCLC's Kilroy Project

Terry Noreault reviewed OCLC's Kilroy Internet database project. OCLC is reaching out to explore the Internet resources at some 800,000 Web sites to establish databases of these resources. Significantly, OCLC is developing statistical means to find these resources and feed them into its traditional Dewey and LC classifications automatically (Scorpion). It is enhancing its classification schema and states that "automatic assignment of classification is feasible." I really don't believe in classification systems for the long run, but apparently, "Classification is dead! Long live Classification!"

Search Engines

Those of you who have been reading Sue Feldman's articles realize what a good communicator she is. Her article in the May 1997 issue of Searcher on search engines is worth reviewing.

In her ASIDIC presentation, Feldman said that searching the Web is good for finding an answer--one good one, that is. If that's what you want, the Web's the place to go. Some may think that's a bit exaggerated, but she wanted to emphasize that that's as far as good searching will go on the Internet. For end users, that's where it's at.

Feldman was particularly good at discussing spamming problems. Other barriers she covered were the size of the Internet, rapid changes of Web pages, and inaccessible text. I liked the fact that she wasn't too enthusiastic about Excite's power-search feature, which was a trend back to Boolean. She stated outright that Boolean is not for Web searching. There are many who naively think that adding Boolean to search engines is a sign

of progress. It isn't for end users.

I don't mean to suggest that Feldman is anti-Web. She isn't. She is for improvement and played a devil's advocate's role--much needed at this time.

Infoseek

Sue Lachance spoke of Infoseek's search engine features opening with "Is it the World Wide Web or the World Wild Web?" The features she discussed were automatic phrase recognition, proper name recognition, distributed search, topical directories created with neural network NET technology, and quality indexing guidelines. She had little to say about distributed search, which is an important new development out of Infoseek. It has received a patent for a method of searching the Web via multiple search engines, a technique that is expected to be fully implemented by the beginning of next year. Infoseek president Steve Kirsch will be addressing this at the Boston meeting in April.

Yes, you've probably guessed it by now. Announcing this Boston meeting is self-serving. I have designed and will chair the program. Please attend anyway. I promise a landmark conference. For more information, contact me or visit the Web site (<http://www.infonortics.com>) and click on "Search Engines Meeting."

The Gorilla Story

Mark Chussil of Advanced Competitive Strategies, Inc. conducts War Games and War Colleges for corporations and other institutions. He is one of my favorite speakers. This was the fourth time I've heard him, and I never tire of his presentations. He was asked to speak because Harry Collier usually designs programs to contain at least one speaker from an allied but remote sphere related to the audience. In this case, we learned a bit about a competitive-intelligence technique.

I paraphrase here his opening story, which teaches us about out-of-the-box thinking, something that's important in these changing times.

In experimenting to see how intelligent a gorilla was, a graduate student shut one up in a room to see how long it would take him to learn to use the doorknob to get out. After a long period of disinterest on the part of the gorilla, the student entered the room to release him, whereupon the gorilla picked up the student and threw him against the wall, thus creating a large hole in the wall through which the gorilla then exited. We learn three lessons from this story: Never assume that there is only one answer to a question, never assume that you know the best answer, and never assume that you are smarter than a gorilla.

After revealing what simulation is all about and what a corporation goes through in applying itself to War Games and the War College, Chussil ended his talk with an Arnold Palmer quote, "The more I practice, the luckier I get."

If you're at all involved in trying to make major competitive changes

and decisions, you should consider Chussil's technique, and you too may get luckier.

Image Retrieval

Gordon Short of Excalibur spoke on advanced techniques in imaging, which are becoming more and more feasible and thus commercial. He spoke of Kanji recognition (recognizing strokes), face recognition (recognizing patterns), scene-change detection (image similarity/difference), and general-image searching ("gestalt," color-shape-texture).

Excalibur seems quite advanced in the commercialization of these techniques. Short showed a picture of a waterlily, which he used as a reference image, and asked his system to retrieve eight similar objects. The eighth likeness was a bunch of bananas, and one could actually detect the seemingly absurd relationship. However, one could limit the search to flowers and come up with a more relevant set.

Concept-based retrieval of images would be the next revolution of image retrieval and Short predicted it was 3-5 years off. (Brenner's law says to double every prediction you hear or see.)

Smart, Dumb, Dumber

David Bellick is search schema manager of MSN Publishing & Tools of the Microsoft Corporation and has been analyzing 2,000 queries he chose randomly from the Web. He started out by saying that NET IR is still inadequate and is technology driven. We all know that, but what he had to say further was evident, although I hadn't realized it--that the end user on the Net today represents the intelligentsia: the people who are likely to have been to college, the big earners who can afford a computer at home as well as at work.

So now I realize we have three levels of users: 1) the smart professionals who know how to search, 2) the smart end users who are pretty dumb at searching, and 3) a mass of uneducated end users who may possibly be yet dumber at searching when they finally begin to use the computer.

Bellick did a KWIC index of the terms used in the 2,000 queries and surprisingly found 4,528 total terms used and only 2,807 of them unique. The top terms were sex (17.5 percent), computer/Internet (14.9 percent), entertainment (14.1 percent), recreation/leisure (12.7 percent), business/investing (5.6 percent), and medicinal/fitness (4.6 percent).

I wonder how intelligent the intelligentsia really are. Think about it.

Ev Brenner managed the Central Abstracting & Indexing Service of the American Petroleum Institute for 30 years and is now a well-known information industry observer. He can be reached by e-mail at 73632.2644@compuserve.com.

COPYRIGHT 1997 Information Today, Inc.

INDUSTRY CODES/NAMES: BUSN Any type of business; LIB Library and Information Science
DESCRIPTORS: Association of Information and Dissemination Centers--Conferences, meetings, seminars, etc.; Information storage and retrieval systems--Product development
PRODUCT/INDUSTRY NAMES: 7375000 (Database Providers)
SIC CODES: 7375 Information retrieval services
FILE SEGMENT: TI File 148

1/9/3 (Item 2 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2001 The Gale Group. All rts. reserv.

09834434 SUPPLIER NUMBER: 19383933 (THIS IS THE FULL TEXT)
Surfing corporate intranets; search tools that control the undertow.
(includes related articles on searching databases via an intranet and
intelligent search agents)
Zorn, Peggy; Emanoil, Mary; Marshall, Lucy; Panek, Mary
Online, v21, n3, p30(16)
May-June, 1997
ISSN: 0146-5422 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 9276 LINE COUNT: 00779

ABSTRACT: The availability of commercial and free intranet search engines promises to meet the unique search and retrieval needs of intranet users. Free and commercial search engine offerings are reviewed based on their technical as well as advanced searching capabilities. Users should choose a search engine based on the type of documents on their site, the site's size, the number of web servers, the server platform and available technical expertise.

TEXT:

Intranets consist of web pages, documents, databases, and other information that sit on a web server or web servers behind an Internet firewall.

THE INTRANET EXPLOSION

The growth in popularity of corporate intranets over the past year has risen to epic proportions. A study by the Forrester Group reveals that over two-thirds of Fortune 500 companies interviewed already have, or are seriously considering implementing a corporate intranet as a means for sharing information across their organizations (1). Exactly what is an intranet and how does it differ from the Internet?

Intranets are internal corporate networks set up to take advantage of popular Internet communication protocols such as TCP/IP and HTTP, and other Internet tools such as web servers, web browsers, and HTML. While the Internet largely provides public, unrestricted access to its content, intranets strictly control access to content, allowing authorized users only. Intranets consist of web pages, documents, databases, and other information that sit on a web server or web servers behind an Internet firewall. Employees use a standard browser, the same browser they use to access information on the Internet, to search and locate internal information. These web sites are devoted to providing access to internal information to employees, while keeping their content secure from the rest of the Internet community.

Intranets are popular with corporations for many reasons:

- * Intranets can be easier and cheaper to implement than traditional groupware solutions like Lotus Notes. Web server and browser software is inexpensive--in some cases, free--and can easily be loaded and operational on a corporate network running the TCP/IP network protocol within a matter of hours.

- * Intranets are scalable--they can start out small with just a few links or home pages in place, and can then grow easily over time to include a huge variety of information with little or no additional investment in infrastructure.

- * Intranets are built using open standards that allow a variety of PC platforms (Windows for Workgroups, Windows 95, OS/2, Macintosh, UNIX, etc.) to access the same information.

- * Intranets can incorporate access to a variety of document and media types including Adobe Acrobat (PDF) documents, HTML, word processing, spreadsheets, sound, video, and graphics applications.

* Empowering employees to independently locate and use a variety of information ranging from online phone directories to Human Resources information can save time and foster employee satisfaction--two things of value to most organizations.

* Finally, intranets allow corporate users to capitalize on their knowledge of using the Internet by using the same software to locate and access internal information. End-users are becoming increasingly comfortable with the web browser and its hypertext linking as an interface to all types of services and information.

SURFING THE INTRANET?

As the content of intranets increases, so does the need for tools that help users locate the information they're looking for quickly and easily. Typical Internet users would find it very difficult, if not impossible, to locate and return to sites they find useful if they didn't use tools like the bookmarking feature of most web browsers and publicly available indexes and search engines such as Alta Vista, Lycos, Open Text, and Yahoo!

The same principles hold true when applied to locating information on intranets. Even the most careful and organized intranet webmaster will find it tough to come up with an organizational scheme for the enterprise web site that makes sense to all users. In addition, the more content added to intranets, the more levels of organizational hierarchy the user will be required to "drill down" through before locating relevant content.

While some users of the Internet may be willing to spend time "surfing" to locate needed information, corporate intranet users and their employers are much more demanding. Companies do not want their employees using unnecessary time to locate needed information in the increasing sea of documents and data available from corporate intranets. By the same token, employees that may be very patient when trying to find information on the giant Internet will not feel that same patience when trying to locate a specific piece of vital internal information.

ENTER...INTRANET SEARCH ENGINES

To help organizations find solutions for locating information on intranets, several intranet search engines--both freeware and commercial--have been developed to address the unique search and retrieval needs of intranet users and developers. These search engines are designed to crawl and index internal web servers and/or portions of these servers to create custom, searchable indexes of the documents and data housed on the servers. They have some features in common with the very large and very popular general Internet search engines, but they also contain some unique capabilities that set them apart from their Internet search engine counterparts.

While both Internet and intranet search engines provide indexing for basic HTML documents, intranet search engines often also provide indexing for other document formats (PDF, word processing, spreadsheet, graphics, databases, etc.) that may be contained in an intranet web site.

Internet search engines are often measured by their ability to provide access to the largest Web indexes, and retrieval from many Internet searches can be overwhelming. Intranet search engines are usually designed to provide more precise data filtering and retrieval, limiting the amount of information the user is required to sift through. To do this, the actual indexing process of an intranet search engine is probably deeper than its Internet counterpart.

According to a recent article in InfoWorld, "...the companies that can configure their search engines for better relevance in search results will be the winners in the intranet field. That difference will come from how their search engines house information" (2).

Corporate librarians and other information professionals can play an active role in evaluating and recommending the use of such products in their organizations; they need to familiarize themselves with the products available and the issues surrounding their selection, implementation, and use. This article takes an in-depth look at the major free and commercial intranet search engines currently available and analyzes differences in features, cost, ease-of-use, and hardware and software requirements. In addition, we'll take a look at some trends that we see affecting the intranet search engine and web server industry that may influence the availability and functionality of intranet search engines in the future.

Information professionals familiar with the indexing and searching process can lend a lot to the evaluation and implementation of intranet search engines within organizations. In-depth knowledge of searching techniques, including use of controlled vocabulary, Boolean operators, proximity operators, and relevancy ranking, is necessary for evaluating the potential effectiveness of the various intranet search engines available. An understanding of, and experience with standard indexing practices and parameters can also ensure that the data contained in the various indexes built on a corporate intranet will facilitate accurate and efficient data retrieval. As the information in intranets grows, the importance of having powerful, accurate, and comprehensive search tools becomes one of the most important issues facing organizations.

RATING THE SEARCH ENGINES

A wide variety of free and commercial search engines for use with intranets exists; products vary greatly in support for search and retrieval, operating system platforms, web server environments, file formats, and cost. This article provides a detailed analysis of eight search engines that can be used for indexing and retrieving documents and information located on an intranet, including:

- * Alta Vista (both Alta Vista Search INTRANET Private eXtension and Alta Vista Search INTRANET XL Private eXtension)

- * Excite for Web Servers
- * Fulcrum Surfboard
- * Glimpse/Harvest
- * ht://dig
- * Open Text Livelink Search
- * Verity Topic Search
- * ZyIndex Webserver

For each search engine covered, we took an in-depth look at technical functionality, searching features supported, results display, and cost. Technical functionality criteria include:

- * server platforms supported (UNIX, NT, VMS, etc.)
- * scalability (can it index an entire intranet, including multiple web servers, and then also build specific indexes based on directory, file type, etc.?)

- * indexable file types (HTML, PDF, word processing, spreadsheet, distributed databases, etc.)

- * technical support availability (toll-free 800 number, World Wide Web site, email, support contracts)

- * price/licensing across multiple sites/web servers

Advanced searching and results display options were also looked at closely. Features focused on include:

- * Boolean logic, including nesting
- * proximity and phrase searching
- * truncation
- * search set manipulation
- * duplicate detection

- * field searching
- * thesaurus or concept searching
- * file formats supported in results display: (i.e., can documents be displayed in results display in native format or just HTML?)
- * relevancy ranking or results sorting
- * keyword-in-context (KWIC) display

Detailed descriptions of the features and functionality of each product examined follow. They are divided into two categories: 1) free, and 2) commercial. In addition, a chart comparing the features of all the search engines covered in this article is provided (Figure 1).

(Figure 1 ILLUSTRATION OMITTED)

THE FREE (OR ALMOST FREE) SEARCH ENGINES

Excite for Web Servers (<http://corp.excite.com/ews.html>)

Excite for Web Servers (EWS) is a free product from Architext Software Inc. It is based on the Excite search engine available for public use on the Internet. This standalone product supports a variety of server platforms, including UNIX, Windows NT, and Macintosh. Document collections, specified by the local system administrator, contain the information about what is to be indexed. These document collections are comprised of CollectionContents, which are lists or descriptions of files to be indexed, as well as the CollectionIndex, which is the searchable index of these files. EWS currently only indexes HTML and text files, but there are plans to index PDF files in the future.

The initial release of EWS was somewhat limited in its scalability, but this has been corrected in version 1.1 of the software. It is now possible to index files on multiple servers, but there is a limitation on individual file sizes and some question about performance with very large document collections. EWS says that several customers have collections larger than 1GB, but as collections grow, performance is impacted with slower search speeds. It does not appear that EWS can include files outside the local intranet, given the indexing options available for the CollectionContents.

Although EWS is a free product, purchase of a maintenance contract is necessary for technical support of the product. Cost for support at this writing (December 1996) is \$995 per machine per year. This support includes free upgrades, as well as email and phone support. A spidering version is slated for release in the first quarter of 1997. This version will be provided free to customers with current maintenance contracts.

Searching features supported in EWS are fairly basic (Figure 2). The search interface is customizable, with the main search mode as either concept-based searching or keyword searching. A concept search allows you to enter a phrase such as pro-choice vote in Michigan.

(Figure 2 ILLUSTRATION OMITTED)

Results are then returned with a confidence rating. However, some results may appear not to contain the terms entered in the search because the concept search attempts to identify concepts rather than the exact terms entered.

Using advanced statistical methods EWS analyzes documents for relationships between terms, and uses those relationships to identify search concepts. EWS stresses that concept searching does not use a thesaurus, which it feels limits precision in results returned. The keyword search functions similarly to a Boolean AND search. Traditional Boolean searching is offered with the current release of EWS; truncation is provided through automatic stemming. Search set manipulation and field searching are not supported at this time.

Results are displayed by relevancy with an option for viewing a summary of the document. The summaries give a brief abstract of the

document so it is not necessary to access the actual document to determine if it satisfies the query. EWS has a Query by Example (QBE) feature to find more documents like those returned. Other features include subject groupings when broad topics are searched, thereby providing the user with a means to refine the search. There is a facility for duplicate detection.

EWS is currently being used to index several public Web sites including those of Nestle, Adobe Systems, and Bell Industries. The search screen and results display is essentially the same at these sites, with some variation depending on the design of the page. EWS is an excellent choice for indexing small collections of documents on a single web server. Organizations just beginning to implement an intranet or small organizations only wanting to index HTML, text, and PDF documents would find EWS to be a reliable search engine. It seems fairly easy both to implement and to administer, and purchasing the maintenance contract would provide these organizations with added support as their intranet grows.

<http://Dig> (<http://htdig.sdsu.edu/>)

<http://Dig>, developed at San Diego State University, is a free and complete web indexing and searching system for an intranet. This standalone product can cover several different web servers at a site but is restricted to the UNIX server platform. As long as the web servers understand the HTTP 1.0 protocol, the web server will work with [ht://Dig](http://Dig). There is no mention of moving to other server platforms in the various documentation describing [ht://Dig](http://Dig).

The software can index all HTML and ASCII text files; other file types are supposed to be searchable in future versions. An interesting security feature of this free software is its ability to search a protected server when the correct password is given. Despite a lack of advanced searching functionality, this security feature is a plus for corporate intranet use.

Samples of how the search engine works can be found from the [ht://Dig](http://Dig) home page by linking to San Diego State University's home page (<http://www.sdsu.edu/>). Basic and advanced searching screens are shown, as well as specific indexes that search through specialized subsections of the university. These links to actual working databases provide an excellent understanding of [ht://Dig](http://Dig) searching capabilities and show the scalability of the program. It is possible to set the software up to search an entire intranet, or a smaller subsection. In addition to searching examples provided through San Diego State University, the program is available immediately for downloading from the [ht://Dig](http://Dig) home page (<http://htdig.sdsu.edu/>).

Search capabilities of [ht://Dig](http://Dig) are fairly basic. From the advanced search screen, it is possible to specify "match all" (AND), "match any" (OR), or "match Boolean" (which accepts the terms AND and OR as commands, plus allows for nesting using parentheses. There are no options for search set manipulation, field searching, or proximity or phrase searching. Truncation is automatic, with no option for searching the root term only.

From the search results screen, your search terms are highlighted. You have the option for long or short results, with the most relevant terms receiving more stars. The current search strategy is listed at the top of the results page for easy reviewing and refining of your search.

Some other search capabilities include the ability to create a controlled vocabulary list by adding keywords to HTML documents, and the ability to do "fuzzy searching," which provides algorithms for search result enhancements, such as finding synonyms.

System requirements and installation notes are clearly listed from the [ht://Dig](http://Dig) home page. Although the notes are quite complete, knowledge of the UNIX operating system and code compiling is necessary. This system,

although not appearing extremely difficult to install, is not turnkey. There are files to download, directories to configure, and scripts to modify. There is a large configuration file for customization once the software has been installed.

Technical support consists of a newsgroup of users; it seems helpful, and an archive of the newsgroup messages covers several common problems. Additionally, there is detailed online documentation, and an email address for Andrew Scherpbier, one of the creators of [ht://Dig](http://Dig).

Harvest/GlimpseHTTP/ WebGlimpse (<http://glimpse.cs.arizona.edu/webglimpse/>)

Harvest is a collection of UNIX-based Internet tools designed to perform several different tasks, such as gathering, extracting, and replicating Web information. The project that created Harvest is officially over, with funding that ended August 1996, although parts of the software collection continue as commercial ventures or as supported by volunteers. One part of this software group is a searching facility that can be applied to both Internet and intranet use.

The search engine software is called Glimpse; it is available and supported in the forms of Glimpse, GlimpseHTTP, and WebGlimpse. In order to use Glimpse on a web site (whether internal or external), you need either GlimpseHTTP or WebGlimpse. According to the GlimpseHTTP Web site, though, WebGlimpse does a superior job of browsing and searching on a single web page. Additionally, WebGlimpse has the capability of indexing and searching several Web servers at once, which GlimpseHTTP cannot do. This section of the evaluations will focus on WebGlimpse, since it is the most appealing to those considering an intranet search engine.

When installed, WebGlimpse inserts a search box at the bottom of every HTML page specified. The search box can be set to search the entire index or the "neighborhood" of the page. The "neighborhood" is defined by the installer as a certain number of links away from the current page. Both this box and the advanced searching box supports Boolean (AND, OR, and NOT), but it is command-based rather than form-based. Thus, someone trying to find Web pages containing the phrase "Arizona Desert" and the word "Windsurfing" would have to type in `Arizona desert;windsurfing` as a search command.

The advanced searching page also includes options for case-sensitive searching, partial-word searching, and the ability to match misspelled words. In WebGlimpse, only HTML and text pages can be searched. Harvest has more search format capabilities, but as you move away from WebGlimpse, you also move away from a complete, supported product.

In WebGlimpse, there are no options for field or concept searching, proximity searching, detecting duplicates, or manipulating search results. From the advanced screen, it is possible to specify the maximum number of files that you would like to have returned by your search.

Your search results include the title of (and a link to) the URL, the date it was last modified, and the list of all lines that matched the query. This results screen thus produces a modified keyword-in-context (KWIC) display, which is extremely useful in determining the relevancy of your retrieval.

Currently, there are no sample databases to search using WebGlimpse. The developers are in the process of releasing a new version, and the "practice" searching was not yet available at this writing. Still, there is a spot on the WebGlimpse home page for sample searching in the future. The entire source code is available for downloading, as is a series of executables for installing the program.

Other Free Intranet Search Utilities: WAIS, `htgrep`, and SWISH
Other free utilities and search tools are available for use on

intranets. However, most require at least some knowledge of CGI scripting, PERL, and/or another programming language to customize for use at your site. In addition, most were developed for the UNIX platform, using utilities and tools available in the UNIX environment that may or may not be transferable to other platforms. Although many were developed using either the PERL or C programming languages, both of which are available for most platforms, the conversion process to other operating environments can be painful without the right expertise. Often, the complexity of the searching that can be done with these free tools and utilities depends on the programming expertise available at your site.

Probably the most widely recognized and powerful searching utility available is WAIS (Wide Area Information Service). WAIS grew out of a project started by Apple Computers, Thinking Machines, and Dow Jones and became one of the most widely used searching tools in the early days of the Internet. WAIS evolved for use with the Web, and a Web version ([www.wais](http://www.wais.com)) is available. WAIS can support fairly advanced searching features such as Boolean, phrase, field, and proximity searches, as well as truncation. There are many varieties of WAIS now in use in addition to [www.wais](http://www.wais.com), including freeWAIS, Son of Wais, Kid-of-WAIS, and a commercial version available from WAIS, Inc. For more information on WAIS or to download the necessary files to get started, see <http://www.eit.com/software/wwwais/wwwais.html> or http://ls6-www.informatik.uni-dortmund.de/ir/projects/freeWAIS-sf/fwsf_1.html.

Another popular tool used to create searchable indexes on intranets is [htgrep](http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html). [Htgrep](http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html) is a UNIX-based CGI script written in PERL that allows queries to any document accessible to your HTTP or web server on a paragraph-by-paragraph basis. [Htgrep](http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html) allows users to create forms-based HTML files that pass all search parameters specified to the searching script. Most sites using [htgrep](http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html) write their own CGI scripts, adapting [htgrep](http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html) to meet their needs and hard-code searching options such as Boolean, truncation, and case-sensitive searches. A FAQ on [htgrep](http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html) is located at <http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html>.

SWISH (Simple Web Indexing System for Humans) is a C program designed to index directories or individual files (usually in HTML format) and provide a search interface to the index created. SWISH uses a configuration file to specify directories and files to search, stop words, and some other basic parameters. SWISH supports Boolean searching and relevancy ranking of results, but not truncation. As is, SWISH can be executed from a command line interface. To use SWISH with an HTML forms interface, you will need to write a CGI program that acts as a gateway between the SWISH program and passes it the necessary searching parameters. To learn more about SWISH or to download the source code, see <http://www.eit.com/software/swish/swish.html>.

These utilities for creating intranet search engines are only the tip of the iceberg in terms of what is available out on the Internet for creating and customizing search engines for use on intranets. There are many more ways to implement intranet search engines, depending on your needs and willingness to program and customize.

COMMERCIAL INTRANET SEARCH ENGINES

Alta Vista (Alta Vista Search INTRANET Private eXtension and Alta Vista Search INTRANET XL Private eXtension) (<http://altavista.software.digital.com/>)

Alta Vista has developed an intranet search engine that uses the same technology that powers the Alta Vista Search Public Service, the popular Internet search engine developed by Digital Equipment Corporation and released in December 1995. Alta Vista Search INTRANET Private eXtension is available in two versions, PX and XL PX. The PX version is for smaller

machines; pricing starts at \$16,000. XL PX is intended for machines with 2GB or more of memory; pricing starts at \$66,000. The software is currently available for Alpha UNIX or Digital Alpha servers, but a version for Windows NT is expected soon.

Like the public search engine, Alta Vista Search Intranet Private eXtension indexes every word on all the web servers on the intranet, as well as specified Internet sites. Using spider software, the search engine crawls the servers behind the company firewall as well as any external Web sites specified, creating an index of every word. This feature is particularly useful to libraries wanting to provide access to information on selected public Web sites through the intranet search engine. Because the software supports multinational intranets, it is able to index servers in multiple locations.

Alta Vista plans to support a wide variety of formats, but the current release indexes only HTML and text files. Database indexing is available with an add-on product--Alta Vista Search Toolkit. Setup and maintenance of the software is designed to need little administration.

Search features and the search interface are the same as with Alta Vista's Internet search engine: Boolean, proximity and phrase searching, field searching, and search set manipulation are available from a forms-based search screen. The results are displayed in relevancy order and can be displayed in standard, detailed, or compact format.

Because the product is relatively new (November 1996 release), it is hard to know how well it will be received by the corporate community. Given the excellent search features offered on the publicly available search engine, as well as its popularity, it is likely to attract a great deal of deserved attention.

Fulcrum Surfboard (<http://www.fultech.com/>)

Fulcrum Surfboard is available from Fulcrum Technologies of Ottawa, Canada. Surfboard is an add-on to SearchServer, Fulcrum's multiplatform search engine driving several Fulcrum products, including SearchBuilder and Find. Supported server platforms are Windows NT and UNIX with support for most any CGI compatible web server including those available from Netscape and Microsoft. The software incorporates security features that limit access based on current firewall specifications or other security needs. It also has a reporting feature so that vital information can be gathered about employee intranet use.

Suriboard has a distributed search architecture that supports open system standards including the use of Z39.50 standards in its search protocol. Fulcrum intentionally built its product using industry standards so that it is able to operate with a wide range of system components. The index is maintained on web servers using MultiGate, Fulcrum's gateway application that accepts search requests, queries the Surfboard index, and returns the results as HTML documents to the end-user. MultiGate is able to access both local and remote sites within the corporate intranet, as well as public Web sites.

Installation is designed to be simple with GUI-based administration tools and wizards to guide the system administrator through setup. However, knowledge of SQL is necessary for maintenance add support. Cost is \$6,250 per server for Surfboard plus an additional \$5,000 for SearchServer and \$295 per seat. Technical support is available through a purchased support contract. This contract includes email and phone support as well as administration courses.

Surfboard indexes and searches most document formats, including HTML, PDF, MS Office documents, relational and WAIS databases, NetNews, and over 50 other document formats. The index which Fulcrum creates is actually a series of tables that hold document attributes as well as a pointer to the

document. The documents themselves remain in their original location. When a search is submitted, SearchServer queries the tables and returns a list of documents with links indicating their original location.

Advanced searching features include Boolean, phrase searching, truncation, date range searching, structured (field) searching and common language searching (Figure 3). Common language searching allows the user to enter a question: How do I install Fulcrum Surfboard? instead of formulating a Boolean search: install* and fulcrum and surfboard. Surfboard offers a feature called SearchObjects for users to bookmark queries or for administrators to design queries for easy access to commonly requested documents.

The search interface in the demo available from the Fulcrum home page is basic, but design of this interface is customizable as is the results display. Results are displayed in a relevancy ranking with search terms highlighted. Documents are converted to HTML on-the-fly if the original format software is not able to be launched from the Web browser.

Customers of Fulcrum include major players in the information technology field such as Microsoft, CompuServe, and Netscape, as well as a variety of other clients. Because Fulcrum offers several information retrieval products and incorporates industry standards into these products, it is attractive to administrators of corporate information tools, particularly corporate intranets. Its distributed architecture, scalability, and ability to be customized make it an excellent choice for organizations with large document collections in a variety of formats and locations.

Open Text: Livelink (<http://www.opentext.com/>)

Open Text Corporation, located in Waterloo, Ontario, Canada, has developed an intranet suite of products called Livelink Intranet. Livelink Search is the search engine of Livelink Intranet; the other three components that complete the Livelink Intranet family include Livelink Library, Workflow, and Collaboration. Many people may already be familiar with Open Text's presence on the Internet via their Open Text Index on the Web (<http://index.opentext.net/>). Livelink Search uses the same full-text indexing software it uses to search the Internet and includes an option to provide both intranet and Internet searching from its default intranet search screen.

Livelink Search currently supports the following server platforms: Windows NT, SUN Solaris and Sun OS, HP/UX, AIX, SGI, and DEC OSF1. The search engine is scalable and guarantees support of document collections of any size. According to a recent Canadian Newswire release, the new search engine is built to handle tens of gigabytes of information, as opposed to hundreds of megabytes common with other search engines (3).

Livelink Spider is the crawler software that locates the documents on the corporate intranet and external Web sites and locally indexes their full text. Documents from relational databases, flat files, HTML, SGML, and 40 other common office data formats can be indexed. It also has the capacity to index Internet mail files and Internet newsgroups. Livelink Spider can be configured to "crawl" to specific domains, server directories, and file types, and conversely, it can be configured not to crawl to specific domains or server directories. Livelink Search has the flexibility to support multiple indexes on one server and multiple indexes on multiple servers, making decentralized collections of information easily searchable.

Open Text offers a variety of support for their products, including training courses for end-users, administrators, and developers; and online reference information and user guides. Customer Service Representatives are available to support any questions regarding functionality, use, and

configuration of Open Text products; however, in order to take advantage of this service, you must subscribe separately to their Customer Assistance Program. At the time of this writing, the price for Livelink Search and Livelink Spider is \$12,000 and \$12,500 per server, respectively. Netscape's Commerce Server communications software is also included in the package.

Full Boolean searching (AND, OR, and NOT) is supported by Livelink Search, as is proximity searching (NEAR), advanced similarity searching (find more results like this one), truncation with a wildcard (*), and full-phrase searching (phrases with no stopwords). Keyword searches look for literal matches of the words, and concept searches use a thesaurus to locate related terms. Searches can also be run to query a specific field of a document (Figure 4). If a Search Application Programmable Interface (API) is purchased and developed for Livelink Search, results can be manipulated for further advanced searching options.

Retrieved results are ranked based on an intelligent ranking algorithm and can be viewed in three different formats: simple ASCII, on-the-fly HTML, or in native format. Livelink Search can convert non-HTML documents on-the-fly so that they can be viewed by any web browser. Document summaries, if not originally provided, can be created using an automatic document summary generator. There is also an option to view the keywords in KWIC mode, where the keywords are highlighted and the user can easily see where the keyword(s) occurs in the retrieved document.

In addition, searches can be restricted to query-specific sections of documents, because the search tool does index documents based on tags or database fields. The client interface is customizable to suit the needs of the users' searching preferences.

Verity SEARCH'97 (<http://www.verity.com>)

Verity, founded in 1988, is the developer of the Topic family of search and retrieval tools for the enterprise and the Internet. In the Fall of 1996, Verity relaunched its entire suite of search and retrieval tools under a new name: SEARCH'97. SEARCH'97 is a comprehensive, flexible platform for deploying search applications across the corporation. The Verity indexing format is being used by over 500 companies worldwide. Some of the companies bundling the search engine into their software include: SAP, Lotus Notes, Individual Inc., Adobe Acrobat, Documentum Inc., Xyvision, Netscape servers, Dow Jones, Reuters, and Ziff Davis

SEARCH'97 can index information from virtually any document format that has been used in the last ten years, including relational databases like Informix, Sybase, and ODI. It is also working towards indexing data from data management files such as Lotus Notes, SAP, Informix, and Documentum. Touted as the mechanism to harness the "corporate memory" of an enterprise, SEARCH'97 facilitates the collection, management, and retrieval of information throughout a corporation and specified sites on the Internet, and makes the data available at an employee's desktop.

The SEARCH'97 platform includes a variety of components: SEARCH '97 Personal, Information Server, Agent Server, Advanced Search and Query enhancements, and Knowledge application tools and advanced navigation. SEARCH'97 Personal is an interface used to initiate search queries, access search agents, and implement searches. SEARCH'97 Personal can locally index Internet Web sites at the individual's computer so that personal Internet Web sites can also be queried along with remote corporate indexes. It is supported from within a web browser or Microsoft Exchange. Results can be viewed in virtually any file type, even if the native application is not available locally. SEARCH'97 Personal is available for UNIX, Mac, Windows 95 and NT.

At the center of the SEARCH'97 framework is the Information Server. The Information Server indexes and manages corporate information--the

"corporate memory"--and uses a web browser or SEARCH'97 Personal as an interface. A web spider is also included to add corporate data and/or Internet sites automatically to the main index.

The full text of documents is indexed; the indexes are updated automatically when data is added, changed, or deleted. The indexing tools support access to virtually any document format including common office document formats, HTML, PDF, and ASCII text. Remote indexing is available to store information from different sites throughout a corporate intranet. The Information Server also acts as the integration point for advanced searching components such as the agent server, enhanced query, visualization, and knowledge and navigation tools. According to Verity's Product Brief, the following platforms support Information Server: Solaris, IBM AIX, HP/UX, Windows NT, DEC Win Alpha, and DEC UNIX.

SEARCH'97 Agent Server automates the search and retrieval process for the individual or corporation. A search profile is prepared and the agent notifies the requester when information or data match the search profile. Individuals customize their information profiles with a set of keywords, specific sources (Internet or intranet sites, including databases) which the query will be run against, and the preferred method for notification (via email, web page, or pager). The agents run continuously and instantaneously alert the user when new information has been added to any of the sources specified in the search profile. Hundreds of thousands of agent profiles can be initiated per server. SEARCH'97 Agent Server currently operates on Solaris 2.5 and Windows NT 3.5.1 platforms. General availability for Agent Server is scheduled for first quarter of 1997 with pricing at approximately \$70,000.

A Technical Support site ([http:// www.verity.com/tech-support/index.html](http://www.verity.com/tech-support/index.html)) is available from the Verity home page. This site includes a technical support information sheet providing phone and fax numbers as well as email addresses for Verity offices worldwide. The technical support information sheet also details the procedures for obtaining technical support. Online information is searchable and includes FAQs, selected data from their Help Desk database, and selected technical notes. Verity also offers a number of educational courses (<http://www.verity.com/educ/index.html>) for their products. Courses are taught at Verity Training Centers in Sunnyvale, CA and Fairfax, VA or can be conducted on-site.

The Verity search engine offers both literal text and Boolean searching capabilities. Other searching options are customizable using standard web forms.

For literal text queries, commas placed between key terms will search on any of those keywords (implied OR). Truncation of words occurs automatically; however, a specific word or phrase can be searched simply by placing quotation marks around the word or phrase. A wildcard can also be used to find variant letters at the beginning of a word or letters. Field searching is available for querying a specific date or author. Proximity operators are also supported; search terms can be specified to show up "near" each other, in the same phrase, sentence, or paragraph. A thesaurus is available to retrieve synonyms for additional search terms.

Natural language queries and query by example (find me more like...) are also supported. The search engine takes the user's query, whether literal or Boolean, and supplements it with "fuzzy logic"--an operator that calculates a "more the better" score to determine relevancy ranking.

Multiple collections (or indexes) can be searched simultaneously by the individual and are selected from a list or drop-down menu. The user can also determine the number of results returned from the search query.

Documents returned are given a score and listed in order of

relevance. Results can be previewed using a rich-text translator, or displayed in the "native format" that can range from data in an Oracle database to Lotus Notes documents to Adobe Acrobat PDF files. Native formats can only be viewed when the requested file format is available locally on the user's computer, or when a suitable viewer is used.

Additional add-on components to the basic SEARCH'97 can increase the flexibility of searching and improve the relevance of results. These optional intelligent search components include: Enhanced Query, Visualization, and Knowledge and Navigation Tools. Enhanced Query uses query technologies such as natural language processing (NLP) and query by example (QBE). A user can type in a search in the form of a question and then use NLP to locate information based on the phrases that were entered. In using QBE, a searcher can copy an example of relevant text from a retrieved result and paste that text in the search form. The QBE engine will then reformulate the search and locate information relevant to the text that was submitted.

The Visualization components (clustering and summarization) make it easier for users to identify relevant information. Clustering organizes the retrieved results into groups based on commonality of terms. The Summarization component creates an overview of individual documents based on an algorithm that determines the significance of the sentences that make up the documents. These summaries are more sophisticated than the typical summaries created by just the document title and following few lines of text.

Navigation tools allow the user to move through documents more easily by using hyperlinks from one document to another. To further facilitate knowledge transfer within an organization, a systems administrator can use Verity's Knowledge Tools. These tools allow administrators to create their own knowledge bases specific to their business environment that include, but also extend beyond the typical functionality of dictionaries and thesauri. These navigation tools would provide more precise search results by filtering out and eliminating irrelevant documents.

ZyIndex Webserver (<http://www.zylab.com/>, <http://www.zylab.nl/>)

ZyLab International, Inc. was founded in 1983 with the introduction of PC-based full-text indexing and retrieval software. Today, ZyLab offers complete web-based publishing and indexing solutions in its ZyIndex Webserver and ZyImage Webserver product lines. ZyIndex Webserver, the software package we will be focusing our attention on in this section, provides full-text indexing of document collections in over 30 formats and makes them searchable through the Internet or corporate intranet. ZyImage Webserver, a companion product to ZyIndex Webserver, combines the ZyImage scanning interface for OCR (optical character recognition) of documents in electronic format with the powerful indexing and search and retrieval engine of ZyIndex Webserver.

All technical specifications and searching functionality described apply to both ZyIndex and ZyImage Webserver. The main difference between the two products is that the ZyImage Webserver offers users the additional benefit of being able to view images of scanned documents with an easy-to-use scanning interface. ZyIndex Webserver sells for \$5,995 complete, while ZyImage Webserver, which includes the ZyImage OCR and ZyIndex software, sells for \$11,200 (price includes annual update service). Both products are licensed to cover a total intranet site. ZyLabs offers a full range of technical support options including an 800 number, electronic mail, Web site, and support contracts.

ZyIndex Webserver supports the most popular HTTP servers, i.e., Microsoft and Netscape. However, ZyIndex Webserver can be used with any existing web server product that is HTTP 1.0 compliant, running on the

Windows NT platform. ZyIndex Webserver provides a proprietary API that handles the search and retrieval process and interfaces with the document index created during the configuration process. In addition, ZyIndex Webserver comes with a set of HTML templates designed to function as the search forms used by end-users through their web browsers, and as the default display format for return and viewing of search results. These templates can be customized to meet client needs.

The ZyIndex Webserver allows clients a great deal of flexibility in indexing features and document and index security. ZyIndex can index document collections located anywhere on the corporate network and can support more than 30 native file formats including all major word processing programs (Word, WordPerfect, etc.), group 4 TIFF, popular database formats (dBase 3 and 4, FoxPro, etc), Lotus, Excel, EPS (encapsulated Postscript), and ASCII and HTML files. However, Adobe Acrobat PDF and Microsoft PowerPoint file formats are not supported at this time.

ZyIndex builds an index based on the documents specified and does not use the documents themselves for retrieval. However, as documents are added or changed, the index is automatically updated. Indexes created by ZyIndex can be very large--up to ten gigabytes can be indexed, or the equivalent of 100 gigabytes of documents. (Indexes of ten gigabytes normally represent approximately 100 gigabytes worth of documents.) If you need to restrict access to certain documents or indexes, ZyIndex allows you to define users and passwords that can be assigned to specific documents, groups of documents, or indexes in order to control security.

Because all indexing done by ZyIndex is on the complete text of the document, and indexes can be very large, a strong set of searching features is needed to ensure accuracy and relevancy of retrieval. ZyIndex Webserver supports Boolean operators and full nesting, phrase searching, advanced proximity searching and truncation, "fuzzy" searches that retrieve words similar to those specified, a "vocabulary" or browse index feature, field searching, and thesaurus for location of synonyms. Searching for numbers or number ranges is supported using standard math operators such as (is less than), (is greater than), =, etc.

In addition, the Concept feature allows web site managers to define searches that cover a particular subject contained in the index, name and save the search strategy, and then display the stored Concept searches for use by end-users searching the index. All of these features are included in an easy-to-use HTML template included with the package.

Retrieved documents, ranked according to relevancy, are automatically translated to HTML on-the-fly for viewing through web browsers regardless of native format, and can then be viewed in native format by launching the appropriate application program. Search terms are highlighted within the context of retrieved documents and users can move "hit to hit" to each occurrence of the term(s) specified in their search request. Another nice feature of the ZyImage Webserver is the ability to view TIFF images directly through the web browser without the use of a helper application or plug-in, by using a TIFF to GIF converter included with the product.

From the ZyLab home page, you can view a demo of ZyIndex Webserver in action on a test database provided by the National Library of Medicine, as well as use a test database set up by ZyLabs to demonstrate a basic installation using the default searching interface and features.

CHOOSING THE RIGHT INTRANET SEARCH ENGINE FOR THE JOB

As demonstrated by the products reviewed in this article, there are many different things that need to be taken into consideration when evaluating and selecting a search engine for an intranet. Size of the site, the type of documents included, the number of web servers, server platform, and technical expertise available are all major factors influencing the

selection of an intranet search engine.

If the intranet site is small and does not contain documents in formats other than HTML and ASCII text, the freeware search engines may be enough to do the job. The frequent downside to these free tools, however, is that advanced technical knowledge is needed to configure and customize the software for site-specific use, and that advanced searching functionality found in the commercial engines is not available. In addition, little formal technical support is offered by any of the free intranet search engines, except for Excite, which charges for its support and maintenance contract.

For large, highly-developed intranet sites, spending the money on a investment. Having the ability to index documents in a variety of file types, including distributed relational databases, and from a variety of locations, both internal and external, makes integrating and then retrieving information on an intranet much easier. Advanced searching features such as field, proximity, and concept searching, as well as the intelligent alerting capabilities promised with the next release of Verity's SEARCH'97 Agent Server, can reduce the number of irrelevant hits produced by a search of a large document collection and automate the search process so that users are automatically notified when content that matches their search profile is added.

WHAT WILL THE FUTURE HOLD?

In this fast-paced, ever-changing world of web-based information retrieval, there are several trends that promise to have a large effect on search and retrieval functionality of intranets.

Not Just Documents

Integration of access to distributed databases (not just documents) with intranet search engines is of paramount importance if intranets are to evolve to the next level of importance in the enterprise.

There are a host of vendors that provide gateways and development tools that can make access to distributed databases from the World Wide Web a reality. Intranet search engine vendors such as Fulcrum, Verity, and Open Text are poised to move to that next level, and may emerge as the favorites for intranet search engines in the near future.

Bundling With Server Software

More and more, web server software packages designed for intranet use are coming bundled with search engines designed to work with the web server software. The two major commercial web server vendors, Netscape and Microsoft, have already capitalized on this trend by including search engines as part of their web server offerings.

Netscape's Enterprise Server comes with the option of purchasing Verity's search engine and using Netscape's Catalog Server (based on Harvest) for indexing document collections. Microsoft's Internet Information Server provides searching functionality through the Microsoft Index Server, a free package that can index and search HTML and file formats created by the software packages in the Microsoft Office Suite. Although both Netscape's Enterprise Server and Microsoft's Internet Information Server provide a built-in searching solution, other search engine products can be used with these web servers, if desired. As intranets grow, it is likely that even though a basic search engine may be included with a web server product, a separate search engine may be purchased as well, depending on the size and complexity of the intranet site.

Intelligent Agents on Alert

The addition of intelligent agents that can "remember" a search query and run it unattended against both internal and external indexes is another emerging trend that will surely become a favorite with intranet users.

Products such as Verity's SEARCH'97 Agent Server and other products mentioned in the sidebar automate the search process and provide vital alerting services that can keep users up-to-date on topics in their areas of interest in "real-time" fashion. The personal search agent products also have the potential, when used with an intranet search engine, to combine results from the inside intranet world with the outside Internet world, giving users a comprehensive view of very current information on specified topics.

Getting It All

Will we ever really be able or even want to search all internal information and external information using one package? Already, typical end-users are becoming frustrated with the amount of retrieval returned by the popular Internet search engines. Intranets, as they grow, have the potential to inspire that same frustration if proper indexing and search and retrieval tools are not developed and implemented.

Gaining balance between providing relevancy, comprehensiveness, and manageability of information on intranets and the Internet as a whole, through development of a set of end-user tools for retrieving and filtering large sets of information, will provide one of the greatest challenges to information professionals in the coming months and years.

REFERENCES

(1) Levitt, Lee. "Intranets: Internet Technologies Deployed Behind the Firewall for Corporate Productivity." Prepared for the Internet Society INET '96 Annual Meeting. <<http://www.process.com/intranets/wp2.htm>>

(2) Balderston, Jim. "Search Engine Vendors Eye Intranets: Intranets Mean That Search Tools Must be Fine-Tuned for Corporate Needs." InfoWorld 18, No. 27 (July 1, 1996): pp. 41.

(3) Canada NewsWire. November 20, 1996.

Communications to the authors should be addressed to Peggy Zorn, Parke-Davis Pharmaceutical Research, 2800 Plymouth Road, Ann Arbor, MI 48150; 313/996-7202; zornm@aa.wl.com and/or Mary Emanoil, Parke-Davis Pharmaceutical Research, 2800 Plymouth Road, Ann Arbor, MI 48150; 313/996-1814; emanoim@aa.wl.com and/or Lucy Marshall, Edge Information Services, 2642 E. Cholla St., Phoenix, AZ 85028; 602/485-9363; edgeinfo@dancris.com and/or Mary Panek, United Technologies Research Center, 411 Silver Lane, MIS 129-01, East Hartford, CT 06108; 860/610-7478; panekmt@utcc.etc.com.

RELATED ARTICLE: The Next Level

SEARCHING DATABASES THROUGH AN INTRANET

While intranet search engines can index and search collections of documents on an intranet, what about including existing databases that a company might have, such as Oracle, Sybase, or Microsoft SQL Server databases? Often, the bulk of the most important information a company has is stored in one of these formats. Gaining access to this vital information from a corporate intranet is a hot issue and will likely form the next wave of major intranet expansion.

Most intranet search engines are not yet able to integrate the searching of information to this next level: tapping both documents and collections of information stored in traditional database formats. Almost all major database vendors have created web-based interfaces and gateways to their products. There are many generic products available that allow you to interface to ODBC (open database connectivity)-compliant databases that use one development environment.

Below is a listing of major vendors competing in the expanding market of web-based connectivity to ODBC-compliant databases, with URLs for more information. Some, such as Oracle WebSystem and Sybase's NetImpact Dynamo, are targeted to specific database products. However, despite these

specializations, all products still claim to have the ability to interface with any ODBC-compliant database product.

ColdFusion <http://www.allaire.com>
Everyware's Tango Enterprise <http://www.everyware.com/>
MEGASOFT Web Transporter <http://www.megasoft.com/>
Microsoft Internet Information Server with Microsoft dbWeb
<http://www.microsoft.com>

Netscape LiveWire <http://www.netscape.com>
NeXT WebObjects <http://www.next.com>
Oracle WebSystem <http://www.oracle.com.sg>
O'Reilly's WebSite Professional <http://www.ora.com/>
Sybase NetImpact Dynamo and web.sql <http://www.sybase.com>
WebDynamics Spider <http://www.w3spider.com>

RELATED ARTICLE: Intelligent Search Agents

Intelligent search agents allow users to create profiles based on their information needs and to simultaneously search selected sites from the external Web, corporate intranet, newsgroups, etc. for the desired information. It is similar to the use of alerting services or SDIs in traditional online searching, except that the intelligent agent can learn from the results, thereby refining the query and returning more valuable information with each new search.

The degree to which intelligent agents are being used varies among software products. Some are simply monitoring tools to alert users when changes have been made to bookmarked sites, but others make associations between search terms and other frequently occurring terms found in search results and then alert the user to these associations. Regardless of the level of agent sophistication, one can expect that software developers will continue to incorporate and improve upon this technology in their products.

Search software that currently uses agent technology includes CyberSearch and WebCompass. Frontier Technologies has announced the release of the 3.0 version of CyberSearch

(<http://www.frontiertech.com/products/cyberseb/csspecl.htm>), its Internet searching and bookmarking utility. Frontier calls the new version of CyberSearch "a global information management tool" because it searches documents on the Internet, intranet, and local PC. Through the use of standard Internet search engines such as Alta Vista, Lycos, Excite, and InfoSeek and server-side indexing of internal documents, this product incorporates the concept of seamless searching among all the information sources accessible to a user.

Quarterdeck Corporation intends to develop a version of its well-reviewed WebCompass software (<http://www.quarterdeck.com/qdeck/products/webcompass/>) that will not only allow users to query multiple search engines, as is the case with its current release, but will also allow for the inclusion of intranet resources. The current version of WebCompass searches multiple Internet search engines simultaneously, sorts the results, and removes duplicate hits. Results are returned in a Microsoft Access database for easy manipulation.

Other software tools that incorporate intelligent agents and that may be beneficial for multisite searching are available. To keep abreast of new developments in the use of intelligent agents for intranet/Internet searching, visit the Complete Intranet Resource (<http://www.intrack.com/intranet/>). This site provides detailed information about intranets, including a list of software sources.

COPYRIGHT 1997 Online Inc.

SPECIAL FEATURES: table; illustration

INDUSTRY CODES/NAMES: BUSN Any type of business; LIB Library and
Information Science
DESCRIPTORS: Intranets--Usage; Online searching--Usage
PRODUCT/INDUSTRY NAMES: 7399200 (Info Services ex Database)
SIC CODES: 7389 Business services, not elsewhere classified
FILE SEGMENT: TI File 148

1/9/4 (Item 1 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2001 The Gale group. All rts. reserv.

05071932 SUPPLIER NUMBER: 19952087 (THIS IS THE FULL TEXT)
The ASIDIC 1997 fall meeting; speakers focused on search-and-retrieval
technologies and techniques.(Association of Information and Dissemination
Centers)
Brenner, Ev
Information Today, v14, n10, p15(2)
Nov, 1997
ISSN: 8755-6286 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 2080 LINE COUNT: 00167

TEXT:

The fall meeting of the Association of Information & Dissemination
Centers (ASIDIC) was held in Seattle September 21-23. The theme for the
one-and-a-half-day meeting, designed and chaired by Harry Collier of
Infonortics, Ltd., was "Incorporating Intelligence into Networked
Information." The conference provided a look at some controversial topics
presented by fine speakers.

NewsNet Keynoter

What a coup of a keynoter! The printed copy of the speakers'
biographies stated, "Andrew Elston is currently evaluating opportunities to
continue his career in publishing and information services while he
oversees the final closing of operations at NewsNet, Inc. this month."

It appeared that Elston was at ASIDIC to tell us why he thinks
NewsNet failed. NewsNet was a 15-year-old, established online database of
about 1,000 newsletters and other news formats. It spent a lot of money
building interfaces to the Web and went live there just 2 years ago. But
when it became clear that NewsNet was no longer a competitive product, the
parent company first tried to sell it, then simply gave it up.

One important factor was that what appeared on the Web was the same
proprietary product as its online version. NewsNet acquired new users on
the Net, but the kind that didn't stick around after retrieving a--meaning
one--quick answer. The traditional users who migrated from online to the
Web and ordinarily stayed online to obtain an average of 5 documents were
now getting an average of 1.2 documents and flitting to other news sites,
many of which delivered the same information for free. Since there are many
competing sites for news--much of it free--that competition simply began to
kill off NewsNet.

Elston rued that NewsNet was the victim of literally not
understanding that the behavior of Web users would be very different from
that of online users. He noted, in his understandably pessimistic mood,
that Knight-Ridder cut off DIALOG, which became Knight-Ridder Information
Services, then hung it out to dry, and has since been trying to sell it.

Elston predicted that DIALOG will break up. I also heard this from a
respected colleague who predicted the break-up in about 5 years.

Let's Get Some Perspective

But, wait up. I'm not quite sure I go along with these predictions on the basis of the NewsNet experience. I brought my love of historical perspective to bear and thought about what advantages online databases had over printed reference texts back in the 1960s when online was considered revolutionary. Online solved important interdisciplinary problems in the sense that good answers could be obtained from a variety of interdisciplinary reference works--and fast.

Many of the databases had controlled indexing, and skilled professionals could search successfully through Boolean and its sophistications.

That was the value-added aspect lacking in the NewsNet product in its shift to the Web. It had nothing more to offer than its text, so it became just another unindexed site, even though it was probably better organized than most of the free sites. It is understandable that it failed.

But that doesn't mean DIALOG and its databases will fail, fall, or crack up that fast. Databases within DIALOG will fail if they have nothing more to offer than text with no real intelligent added-value features. Will that kill off DIALOG? It depends on whether downsizing in corporate information centers continues unabated. There are indications that, after the downsizing onslaught, there may be a new middle road to the role of intermediary.

Elston spoke of the resurgence of newspapers in print in the 1990s. The recent move of The New York Times in dividing its present version into more sections seems germane. Another interesting point made by Elston concerned venture capital. The venturers are not looking at established companies but to those with uncharted futures. The young, untried technology bucks are looking good. The venturers are now true adventurers.

In a summary wrap-up at the end of a conference on search engines a couple of years ago, one point made by James Callam of the University of Massachusetts was that "Boolean is dead! Long live Boolean!" Recent conferences, including this one, have led me to wonder about "Online is dead! Long live Online!," even though the NewsNet demise was a rather frighteningly speedy one.

IsoQuest

I was intrigued by IsoQuest and its Data Extraction Technology Tool-Kit. Tony Hall reviewed how far a 15-month-old software company has come with its NetOwl (see related news announcement in the Internet Publishing Today section) He spoke of an automatic-index software tool that can browse dynamically for company names, place names, and people names, and also can automatically extract pieces of text from full text, thus providing a useful summarization of that article.

Automatic abstracting rears its head once more. Individual, Excalibur, and Infoseek are three retrieval companies that have bought into this product. Entity extraction, the more scholarly name for this kind of retrieval, will be the subject of a major talk presented by the president of IsoQuest, Paul Jacobs, at the "Search Engine and Beyond" conference to be held in Boston April 1-2, 1998.

OCLC's Kilroy Project

Terry Noreault reviewed OCLC's Kilroy Internet database project. OCLC is reaching out to explore the Internet resources at some 800,000 Web sites to establish databases of these resources. Significantly, OCLC is developing statistical means to find these resources and feed them into its traditional Dewey and LC classifications automatically (Scorpion). It is enhancing its classification schema and states that "automatic assignment of classification is feasible." I really don't believe in classification systems for the long run, but apparently, "Classification is dead! Long

live Classification!"

Search Engines

Those of you who have been reading Sue Feldman's articles realize what a good communicator she is. Her article in the May 1997 issue of Searcher on search engines is worth reviewing.

In her ASIDIC presentation, Feldman said that searching the Web is good for finding an answer--one good one, that is. If that's what you want, the Web's the place to go. Some may think that's a bit exaggerated, but she wanted to emphasize that that's as far as good searching will go on the Internet. For end users, that's where it's at.

Feldman was particularly good at discussing spamming problems. Other barriers she covered were the size of the Internet, rapid changes of Web pages, and inaccessible text. I liked the fact that she wasn't too enthusiastic about Excite's power-search feature, which was a trend back to Boolean. She stated outright that Boolean is not for Web searching. There are many who naively think that adding Boolean to search engines is a sign of progress. It isn't for end users.

I don't mean to suggest that Feldman is anti-Web. She isn't. She is for improvement and played a devil's advocate's role--much needed at this time.

Infoseek

Sue Lachance spoke of Infoseek's search engine features opening with "Is it the World Wide Web or the World Wild Web?" The features she discussed were automatic phrase recognition, proper name recognition, distributed search, topical directories created with neural network NET technology, and quality indexing guidelines. She had little to say about distributed search, which is an important new development out of Infoseek. It has received a patent for a method of searching the Web via multiple search engines, a technique that is expected to be fully implemented by the beginning of next year. Infoseek president Steve Kirsch will be addressing this at the Boston meeting in April.

Yes, you've probably guessed it by now. Announcing this Boston meeting is self-serving. I have designed and will chair the program. Please attend anyway. I promise a landmark conference. For more information, contact me or visit the Web site (<http://www.infonortics.com>) and click on "Search Engines Meeting."

The Gorilla Story

Mark Chussil of Advanced Competitive Strategies, Inc. conducts War Games and War Colleges for corporations and other institutions. He is one of my favorite speakers. This was the fourth time I've heard him, and I never tire of his presentations. He was asked to speak because Harry Collier usually designs programs to contain at least one speaker from an allied but remote sphere related to the audience. In this case, we learned a bit about a competitive-intelligence technique.

I paraphrase here his opening story, which teaches us about out-of-the-box thinking, something that's important in these changing times.

In experimenting to see how intelligent a gorilla was, a graduate student shut one up in a room to see how long it would take him to learn to use the doorknob to get out. After a long period of disinterest on the part of the gorilla, the student entered the room to release him, whereupon the gorilla picked up

the student and threw him
against the wall, thus creating a
large hole in the wall through
which the gorilla then exited.
We learn three lessons from
this story: Never assume that
there is only one answer to a
question, never assume that
you know the best answer, and
never assume that you are
smarter than a gorilla.

After revealing what simulation is all about and what a corporation goes through in applying itself to War Games and the War College, Chussil ended his talk with an Arnold Palmer quote, "The more I practice, the luckier I get."

If you're at all involved in trying to make major competitive changes and decisions, you should consider Chussil's technique, and you too may get luckier.

Image Retrieval

Gordon Short of Excalibur spoke on advanced techniques in imaging, which are becoming more and more feasible and thus commercial. He spoke of Kanji recognition (recognizing strokes), face recognition (recognizing patterns), scene-change detection (image similarity/difference), and general-image searching ("gestalt," color-shape-texture).

Excalibur seems quite advanced in the commercialization of these techniques. Short showed a picture of a waterlily, which he used as a reference image, and asked his system to retrieve eight similar objects. The eighth likeness was a bunch of bananas, and one could actually detect the seemingly absurd relationship. However, one could limit the search to flowers and come up with a more relevant set.

Concept-based retrieval of images would be the next revolution of image retrieval and Short predicted it was 3-5 years off. (Brenner's law says to double every prediction you hear or see.)

Smart, Dumb, Dumber

David Bellick is search schema manager of MSN Publishing & Tools of the Microsoft Corporation and has been analyzing 2,000 queries he chose randomly from the Web. He started out by saying that NET IR is still inadequate and is technology driven. We all know that, but what he had to say further was evident, although I hadn't realized it--that the end user on the Net today represents the intelligentsia: the people who are likely to have been to college, the big earners who can afford a computer at home as well as at work.

So now I realize we have three levels of users: 1) the smart professionals who know how to search, 2) the smart end users who are pretty dumb at searching, and 3) a mass of uneducated end users who may possibly be yet dumber at searching when they finally begin to use the computer.

Bellick did a KWIC index of the terms used in the 2,000 queries and surprisingly found 4,528 total terms used and only 2,807 of them unique. The top terms were sex (17.5 percent), computer/Internet (14.9 percent), entertainment (14.1 percent), recreation/leisure (12.7 percent), business/investing (5.6 percent), and medicinal/fitness (4.6 percent).

I wonder how intelligent the intelligentsia really are. Think about it.

Ev Brenner managed the Central Abstracting & Indexing Service of the American Petroleum Institute for 30 years and is now a well-known information industry observer. He can be reached by e-mail at 73632.2644@compuserve.com.

COPYRIGHT 1997 Information Today, Inc.

DESCRIPTORS: Association of Information and Dissemination Centers--
Conferences, meetings, seminars, etc.; Information storage and retrieval
systems--Product development
PRODUCT/INDUSTRY NAMES: 7375000 (Database Providers)
SIC CODES: 7375 Information retrieval services
FILE SEGMENT: TI File 148

1/9/5 (Item 2 from file: 47)
DIALOG(R) File 47:Gale Group Magazine DB(TM)
(c) 2001 The Gale group. All rts. reserv.

04830012 SUPPLIER NUMBER: 19383933 (THIS IS THE FULL TEXT)
Surfing corporate intranets; search tools that control the undertow.
(includes related articles on searching databases via an intranet and
intelligent search agents)
Zorn, Peggy; Emanoil, Mary; Marshall, Lucy; Panek, Mary
Online, v21, n3, p30(16)
May-June, 1997
ISSN: 0146-5422 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 9276 LINE COUNT: 00779

ABSTRACT: The availability of commercial and free intranet search engines
promises to meet the unique search and retrieval needs of intranet users.
Free and commercial search engine offerings are reviewed based on their
technical as well as advanced searching capabilities. Users should choose a
search engine based on the type of documents on their site, the site's
size, the number of web servers, the server platform and available
technical expertise.

TEXT:

Intranets consist of web pages, documents, databases, and other
information that sit on a web server or web servers behind an Internet
firewall.

THE INTRANET EXPLOSION

The growth in popularity of corporate intranets over the past year
has risen to epic proportions. A study by the Forrester Group reveals that
over two-thirds of Fortune 500 companies interviewed already have, or are
seriously considering implementing a corporate intranet as a means for
sharing information across their organizations (1). Exactly what is an
intranet and how does it differ from the Internet?

Intranets are internal corporate networks set up to take advantage of
popular Internet communication protocols such as TCP/IP and HTTP, and other
Internet tools such as web servers, web browsers, and HTML. While the
Internet largely provides public, unrestricted access to its content,
intranets strictly control access to content, allowing authorized users
only. Intranets consist of web pages, documents, databases, and other
information that sit on a web server or web servers behind an Internet
firewall. Employees use a standard browser, the same browser they use to
access information on the Internet, to search and locate internal
information. These web sites are devoted to providing access to internal
information to employees, while keeping their content secure from the rest
of the Internet community.

Intranets are popular with corporations for many reasons:

* Intranets can be easier and cheaper to implement than traditional

groupware solutions like Lotus Notes. Web server and browser software is inexpensive--in some cases, free--and can easily be loaded and operational on a corporate network running the TCP/IP network protocol within a matter of hours.

- * Intranets are scalable--they can start out small with just a few links or home pages in place, and can then grow easily over time to include a huge variety of information with little or no additional investment in infrastructure.

- * Intranets are built using open standards that allow a variety of PC platforms (Windows for Workgroups, Windows 95, OS/2, Macintosh, UNIX, etc.) to access the same information.

- * Intranets can incorporate access to a variety of document and media types including Adobe Acrobat (PDF) documents, HTML, word processing, spreadsheets, sound, video, and graphics applications.

- * Empowering employees to independently locate and use a variety of information ranging from online phone directories to Human Resources information can save time and foster employee satisfaction--two things of value to most organizations.

- * Finally, intranets allow corporate users to capitalize on their knowledge of using the Internet by using the same software to locate and access internal information. End-users are becoming increasingly comfortable with the web browser and its hypertext linking as an interface to all types of services and information.

SURFING THE INTRANET?

As the content of intranets increases, so does the need for tools that help users locate the information they're looking for quickly and easily. Typical Internet users would find it very difficult, if not impossible, to locate and return to sites they find useful if they didn't use tools like the bookmarking feature of most web browsers and publicly available indexes and search engines such as Alta Vista, Lycos, Open Text, and Yahoo!

The same principles hold true when applied to locating information on intranets. Even the most careful and organized intranet webmaster will find it tough to come up with an organizational scheme for the enterprise web site that makes sense to all users. In addition, the more content added to intranets, the more levels of organizational hierarchy the user will be required to "drill down" through before locating relevant content.

While some users of the Internet may be willing to spend time "surfing" to locate needed information, corporate intranet users and their employers are much more demanding. Companies do not want their employees using unnecessary time to locate needed information in the increasing sea of documents and data available from corporate intranets. By the same token, employees that may be very patient when trying to find information on the giant Internet will not feel that same patience when trying to locate a specific piece of vital internal information.

ENTER...INTRANET SEARCH ENGINES

To help organizations find solutions for locating information on intranets, several intranet search engines--both freeware and commercial--have been developed to address the unique search and retrieval needs of intranet users and developers. These search engines are designed to crawl and index internal web servers and/or portions of these servers to create custom, searchable indexes of the documents and data housed on the servers. They have some features in common with the very large and very popular general Internet search engines, but they also contain some unique capabilities that set them apart from their Internet search engine counterparts.

While both Internet and intranet search engines provide indexing for

basic HTML documents, intranet search engines often also provide indexing for other document formats (PDF, word processing, spreadsheet, graphics, databases, etc.) that may be contained in an intranet web site.

Internet search engines are often measured by their ability to provide access to the largest Web indexes, and retrieval from many Internet searches can be overwhelming. Intranet search engines are usually designed to provide more precise data filtering and retrieval, limiting the amount of information the user is required to sift through. To do this, the actual indexing process of an intranet search engine is probably deeper than its Internet counterpart.

According to a recent article in InfoWorld, "...the companies that can configure their search engines for better relevance in search results will be the winners in the intranet field. That difference will come from how their search engines house information" (2).

Corporate librarians and other information professionals can play an active role in evaluating and recommending the use of such products in their organizations; they need to familiarize themselves with the products available and the issues surrounding their selection, implementation, and use. This article takes an in-depth look at the major free and commercial intranet search engines currently available and analyzes differences in features, cost, ease-of-use, and hardware and software requirements. In addition, we'll take a look at some trends that we see affecting the intranet search engine and web server industry that may influence the availability and functionality of intranet search engines in the future.

Information professionals familiar with the indexing and searching process can lend a lot to the evaluation and implementation of intranet search engines within organizations. In-depth knowledge of searching techniques, including use of controlled vocabulary, Boolean operators, proximity operators, and relevancy ranking, is necessary for evaluating the potential effectiveness of the various intranet search engines available. An understanding of, and experience with standard indexing practices and parameters can also ensure that the data contained in the various indexes built on a corporate intranet will facilitate accurate and efficient data retrieval. As the information in intranets grows, the importance of having powerful, accurate, and comprehensive search tools becomes one of the most important issues facing organizations.

RATING THE SEARCH ENGINES

A wide variety of free and commercial search engines for use with intranets exists; products vary greatly in support for search and retrieval, operating system platforms, web server environments, file formats, and cost. This article provides a detailed analysis of eight search engines that can be used for indexing and retrieving documents and information located on an intranet, including:

- * Alta Vista (both Alta Vista Search INTRANET Private eXtension and Alta Vista Search INTRANET XL Private eXtension)

- * Excite for Web Servers
- * Fulcrum Surfboard
- * Glimpse/Harvest
- * ht://dig
- * Open Text Livelink Search
- * Verity Topic Search
- * ZyIndex Webserver

For each search engine covered, we took an in-depth look at technical functionality, searching features supported, results display, and cost. Technical functionality criteria include:

- * server platforms supported (UNIX, NT, VMS, etc.)
- * scalability (can it index an entire intranet, including multiple

web servers, and then also build specific indexes based on directory, file type, etc.?)

- * indexable file types (HTML, PDF, word processing, spreadsheet, distributed databases, etc.)

- * technical support availability (toll-free 800 number, World Wide Web site, email, support contracts)

- * price/licensing across multiple sites/web servers

Advanced searching and results display options were also looked at closely. Features focused on include:

- * Boolean logic, including nesting

- * proximity and phrase searching

- * truncation

- * search set manipulation

- * duplicate detection

- * field searching

- * thesaurus or concept searching

- * file formats supported in results display: (i.e., can documents be displayed in results display in native format or just HTML?)

- * relevancy ranking or results sorting

- * keyword-in-context (KWIC) display

Detailed descriptions of the features and functionality of each product examined follow. They are divided into two categories: 1) free, and 2) commercial. In addition, a chart comparing the features of all the search engines covered in this article is provided (Figure 1).

(Figure 1 ILLUSTRATION OMITTED)

THE FREE (OR ALMOST FREE) SEARCH ENGINES

Excite for Web Servers (<http://corp.excite.com/ews.html>)

Excite for Web Servers (EWS) is a free product from Architext Software Inc. It is based on the Excite search engine available for public use on the Internet. This standalone product supports a variety of server platforms, including UNIX, Windows NT, and Macintosh. Document collections, specified by the local system administrator, contain the information about what is to be indexed. These document collections are comprised of CollectionContents, which are lists or descriptions of files to be indexed, as well as the CollectionIndex, which is the searchable index of these files. EWS currently only indexes HTML and text files, but there are plans to index PDF files in the future.

The initial release of EWS was somewhat limited in its scalability, but this has been corrected in version 1.1 of the software. It is now possible to index files on multiple servers, but there is a limitation on individual file sizes and some question about performance with very large document collections. EWS says that several customers have collections larger than 1GB, but as collections grow, performance is impacted with slower search speeds. It does not appear that EWS can include files outside the local intranet, given the indexing options available for the CollectionContents.

Although EWS is a free product, purchase of a maintenance contract is necessary for technical support of the product. Cost for support at this writing (December 1996) is \$995 per machine per year. This support includes free upgrades, as well as email and phone support. A spidering version is slated for release in the first quarter of 1997. This version will be provided free to customers with current maintenance contracts.

Searching features supported in EWS are fairly basic (Figure 2). The search interface is customizable, with the main search mode as either concept-based searching or keyword searching. A concept search allows you to enter a phrase such as pro-choice vote in Michigan.

(Figure 2 ILLUSTRATION OMITTED)

Results are then returned with a confidence rating. However, some results may appear not to contain the terms entered in the search because the concept search attempts to identify concepts rather than the exact terms entered.

Using advanced statistical methods EWS analyzes documents for relationships between terms, and uses those relationships to identify search concepts. EWS stresses that concept searching does not use a thesaurus, which it feels limits precision in results returned. The keyword search functions similarly to a Boolean AND search. Traditional Boolean searching is offered with the current release of EWS; truncation is provided through automatic stemming. Search set manipulation and field searching are not supported at this time.

Results are displayed by relevancy with an option for viewing a summary of the document. The summaries give a brief abstract of the document so it is not necessary to access the actual document to determine if it satisfies the query. EWS has a Query by Example (QBE) feature to find more documents like those returned. Other features include subject groupings when broad topics are searched, thereby providing the user with a means to refine the search. There is a facility for duplicate detection.

EWS is currently being used to index several public Web sites including those of Nestle, Adobe Systems, and Bell Industries. The search screen and results display is essentially the same at these sites, with some variation depending on the design of the page. EWS is an excellent choice for indexing small collections of documents on a single web server. Organizations just beginning to implement an intranet or small organizations only wanting to index HTML, text, and PDF documents would find EWS to be a reliable search engine. It seems fairly easy both to implement and to administer, and purchasing the maintenance contract would provide these organizations with added support as their intranet grows.

[ht://Dig](http://Dig) (<http://htdig.sdsu.edu/>)

[ht://Dig](http://Dig), developed at San Diego State University, is a free and complete web indexing and searching system for an intranet. This standalone product can cover several different web servers at a site but is restricted to the UNIX server platform. As long as the web servers understand the HTTP 1.0 protocol, the web server will work with [ht://Dig](http://Dig). There is no mention of moving to other server platforms in the various documentation describing [ht://Dig](http://Dig).

The software can index all HTML and ASCII text files; other file types are supposed to be searchable in future versions. An interesting security feature of this free software is its ability to search a protected server when the correct password is given. Despite a lack of advanced searching functionality, this security feature is a plus for corporate intranet use.

Samples of how the search engine works can be found from the [ht://Dig](http://Dig) home page by linking to San Diego State University's home page (<http://www.sdsu.edu/>). Basic and advanced searching screens are shown, as well as specific indexes that search through specialized subsections of the university. These links to actual working databases provide an excellent understanding of [ht://Dig](http://Dig) searching capabilities and show the scalability of the program. It is possible to set the software up to search an entire intranet, or a smaller subsection. In addition to searching examples provided through San Diego State University, the program is available immediately for downloading from the [ht://Dig](http://Dig) home page (<http://htdig.sdsu.edu/>).

Search capabilities of [ht://Dig](http://Dig) are fairly basic. From the advanced search screen, it is possible to specify "match all" (AND), "match any" (OR), or "match Boolean" (which accepts the terms AND and OR as commands,

plus allows for nesting using parentheses. There are no options for search set manipulation, field searching, or proximity or phrase searching. Truncation is automatic, with no option for searching the root term only.

From the search results screen, your search terms are highlighted. You have the option for long or short results, with the most relevant terms receiving more stars. The current search strategy is listed at the top of the results page for easy reviewing and refining of your search.

Some other search capabilities include the ability to create a controlled vocabulary list by adding keywords to HTML documents, and the ability to do "fuzzy searching," which provides algorithms for search result enhancements, such as finding synonyms.

System requirements and installation notes are clearly listed from the [ht://Dig](http://Dig) home page. Although the notes are quite complete, knowledge of the UNIX operating system and code compiling is necessary. This system, although not appearing extremely difficult to install, is not turnkey. There are files to download, directories to configure, and scripts to modify. There is a large configuration file for customization once the software has been installed.

Technical support consists of a newsgroup of users; it seems helpful, and an archive of the newsgroup messages covers several common problems. Additionally, there is detailed online documentation, and an email address for Andrew Scherpbier, one of the creators of [ht://Dig](http://Dig).

Harvest/GlimpseHTTP/ WebGlimpse (<http://glimpse.cs.arizona.edu/webglimpse/>)

Harvest is a collection of UNIX-based Internet tools designed to perform several different tasks, such as gathering, extracting, and replicating Web information. The project that created Harvest is officially over, with funding that ended August 1996, although parts of the software collection continue as commercial ventures or as supported by volunteers. One part of this software group is a searching facility that can be applied to both Internet and intranet use.

The search engine software is called Glimpse; it is available and supported in the forms of Glimpse, GlimpseHTTP, and WebGlimpse. In order to use Glimpse on a web site (whether internal or external), you need either GlimpseHTTP or WebGlimpse. According to the GlimpseHTTP Web site, though, WebGlimpse does a superior job of browsing and searching on a single web page. Additionally, WebGlimpse has the capability of indexing and searching several Web servers at once, which GlimpseHTTP cannot do. This section of the evaluations will focus on WebGlimpse, since it is the most appealing to those considering an intranet search engine.

When installed, WebGlimpse inserts a search box at the bottom of every HTML page specified. The search box can be set to search the entire index or the "neighborhood" of the page. The "neighborhood" is defined by the installer as a certain number of links away from the current page. Both this box and the advanced searching box supports Boolean (AND, OR, and NOT), but it is command-based rather than form-based. Thus, someone trying to find Web pages containing the phrase "Arizona Desert" and the word "Windsurfing" would have to type in `Arizona desert;windsurfing` as a search command.

The advanced searching page also includes options for case-sensitive searching, partial-word searching, and the ability to match misspelled words. In WebGlimpse, only HTML and text pages can be searched. Harvest has more search format capabilities, but as you move away from WebGlimpse, you also move away from a complete, supported product.

In WebGlimpse, there are no options for field or concept searching, proximity searching, detecting duplicates, or manipulating search results. From the advanced screen, it is possible to specify the maximum number of

files that you would like to have returned by your search.

Your search results include the title of (and a link to) the URL, the date it was last modified, and the list of all lines that matched the query. This results screen thus produces a modified keyword-in-context (KWIC) display, which is extremely useful in determining the relevancy of your retrieval.

Currently, there are no sample databases to search using WebGlimpse. The developers are in the process of releasing a new version, and the "practice" searching was not yet available at this writing. Still, there is a spot on the WebGlimpse home page for sample searching in the future. The entire source code is available for downloading, as is a series of executables for installing the program.

Other Free Intranet Search Utilities: WAIS, htgrep, and SWISH

Other free utilities and search tools are available for use on intranets. However, most require at least some knowledge of CGI scripting, PERL, and/or another programming language to customize for use at your site. In addition, most were developed for the UNIX platform, using utilities and tools available in the UNIX environment that may or may not be transferable to other platforms. Although many were developed using either the PERL or C programming languages, both of which are available for most platforms, the conversion process to other operating environments can be painful without the right expertise. Often, the complexity of the searching that can be done with these free tools and utilities depends on the programming expertise available at your site.

Probably the most widely recognized and powerful searching utility available is WAIS (Wide Area Information Service). WAIS grew out of a project started by Apple Computers, Thinking Machines, and Dow Jones and became one of the most widely used searching tools in the early days of the Internet. WAIS evolved for use with the Web, and a Web version ([www.wais](http://www.wais.com)) is available. WAIS can support fairly advanced searching features such as Boolean, phrase, field, and proximity searches, as well as truncation. There are many varieties of WAIS now in use in addition to [www.wais](http://www.wais.com), including freeWAIS, Son of Wais, Kid-of-WAIS, and a commercial version available from WAIS, Inc. For more information on WAIS or to download the necessary files to get started, see <http://www.eit.com/software/www.wais/www.wais.html> or http://ls6-www.informatik.uni-dortmund.de/ir/projects/freeWAIS-sf/fwsf_1.html.

Another popular tool used to create searchable indexes on intranets is htgrep. Htgrep is a UNIX-based CGI script written in PERL that allows queries to any document accessible to your HTTP or web server on a paragraph-by-paragraph basis. Htgrep allows users to create forms-based HTML files that pass all search parameters specified to the searching script. Most sites using htgrep write their own CGI scripts, adapting htgrep to meet their needs and hard-code searching options such as Boolean, truncation, and case-sensitive searches. A FAQ on htgrep is located at <http://iamwww.unibe.ch/~scg/Src/Doc/htgrep.html>.

SWISH (Simple Web Indexing System for Humans) is a C program designed to index directories or individual files (usually in HTML format) and provide a search interface to the index created. SWISH uses a configuration file to specify directories and files to search, stop words, and some other basic parameters. SWISH supports Boolean searching and relevancy ranking of results, but not truncation. As is, SWISH can be executed from a command line interface. To use SWISH with an HTML forms interface, you will need to write a CGI program that acts as a gateway between the SWISH program and passes it the necessary searching parameters. To learn more about SWISH or to download the source code, see <http://www.eit.com/software/swish/swish.html>.

These utilities for creating intranet search engines are only the tip of the iceberg in terms of what is available out on the Internet for creating and customizing search engines for use on intranets. There are many more ways to implement intranet search engines, depending on your needs and willingness to program and customize.

COMMERCIAL INTRANET SEARCH ENGINES

Alta Vista (Alta Vista Search INTRANET Private eXtension and Alta Vista Search INTRANET XL Private eXtension) (<http://altavista.software.digital.com/>)

Alta Vista has developed an intranet search engine that uses the same technology that powers the Alta Vista Search Public Service, the popular Internet search engine developed by Digital Equipment Corporation and released in December 1995. Alta Vista Search INTRANET Private eXtension is available in two versions, PX and XL PX. The PX version is for smaller machines; pricing starts at \$16,000. XL PX is intended for machines with 2GB or more of memory; pricing starts at \$66,000. The software is currently available for Alpha UNIX or Digital Alpha servers, but a version for Windows NT is expected soon.

Like the public search engine, Alta Vista Search Intranet Private eXtension indexes every word on all the web servers on the intranet, as well as specified Internet sites. Using spider software, the search engine crawls the servers behind the company firewall as well as any external Web sites specified, creating an index of every word. This feature is particularly useful to libraries wanting to provide access to information on selected public Web sites through the intranet search engine. Because the software supports multinational intranets, it is able to index servers in multiple locations.

Alta Vista plans to support a wide variety of formats, but the current release indexes only HTML and text files. Database indexing is available with an add-on product--Alta Vista Search Toolkit. Setup and maintenance of the software is designed to need little administration.

Search features and the search interface are the same as with Alta Vista's Internet search engine: Boolean, proximity and phrase searching, field searching, and search set manipulation are available from a forms-based search screen. The results are displayed in relevancy order and can be displayed in standard, detailed, or compact format.

Because the product is relatively new (November 1996 release), it is hard to know how well it will be received by the corporate community. Given the excellent search features offered on the publicly available search engine, as well as its popularity, it is likely to attract a great deal of deserved attention.

Fulcrum Surfboard (<http://www.fultech.com/>)

Fulcrum Surfboard is available from Fulcrum Technologies of Ottawa, Canada. Surfboard is an add-on to SearchServer, Fulcrum's multiplatform search engine driving several Fulcrum products, including SearchBuilder and Find. Supported server platforms are Windows NT and UNIX with support for most any CGI compatible web server including those available from Netscape and Microsoft. The software incorporates security features that limit access based on current firewall specifications or other security needs. It also has a reporting feature so that vital information can be gathered about employee intranet use.

Surfboard has a distributed search architecture that supports open system standards including the use of Z39.50 standards in its search protocol. Fulcrum intentionally built its product using industry standards so that it is able to operate with a wide range of system components. The index is maintained on web servers using MultiGate, Fulcrum's gateway application that accepts search requests, queries the Surfboard index, and

returns the results as HTML documents to the end-user. MultiGate is able to access both local and remote sites within the corporate intranet, as well as public Web sites.

Installation is designed to be simple with GUI-based administration tools and wizards to guide the system administrator through setup. However, knowledge of SQL is necessary for maintenance and support. Cost is \$6,250 per server for Surfboard plus an additional \$5,000 for SearchServer and \$295 per seat. Technical support is available through a purchased support contract. This contract includes email and phone support as well as administration courses.

Surfboard indexes and searches most document formats, including HTML, PDF, MS Office documents, relational and WAIS databases, NetNews, and over 50 other document formats. The index which Fulcrum creates is actually a series of tables that hold document attributes as well as a pointer to the document. The documents themselves remain in their original location. When a search is submitted, SearchServer queries the tables and returns a list of documents with links indicating their original location.

Advanced searching features include Boolean, phrase searching, truncation, date range searching, structured (field) searching and common language searching (Figure 3). Common language searching allows the user to enter a question: How do I install Fulcrum Surfboard? instead of formulating a Boolean search: install* and fulcrum and surfboard. Surfboard offers a feature called SearchObjects for users to bookmark queries or for administrators to design queries for easy access to commonly requested documents.

The search interface in the demo available from the Fulcrum home page is basic, but design of this interface is customizable as is the results display. Results are displayed in a relevancy ranking with search terms highlighted. Documents are converted to HTML on-the-fly if the original format software is not able to be launched from the Web browser.

Customers of Fulcrum include major players in the information technology field such as Microsoft, CompuServe, and Netscape, as well as a variety of other clients. Because Fulcrum offers several information retrieval products and incorporates industry standards into these products, it is attractive to administrators of corporate information tools, particularly corporate intranets. Its distributed architecture, scalability, and ability to be customized make it an excellent choice for organizations with large document collections in a variety of formats and locations.

Open Text: Livelink (<http://www.opentext.com/>)

Open Text Corporation, located in Waterloo, Ontario, Canada, has developed an intranet suite of products called Livelink Intranet. Livelink Search is the search engine of Livelink Intranet; the other three components that complete the Livelink Intranet family include Livelink Library, Workflow, and Collaboration. Many people may already be familiar with Open Text's presence on the Internet via their Open Text Index on the Web (<http://index.opentext.net/>). Livelink Search uses the same full-text indexing software it uses to search the Internet and includes an option to provide both intranet and Internet searching from its default intranet search screen.

Livelink Search currently supports the following server platforms: Windows NT, SUN Solaris and Sun OS, HP/UX, AIX, SGI, and DEC OSF1. The search engine is scalable and guarantees support of document collections of any size. According to a recent Canadian Newswire release, the new search engine is built to handle tens of gigabytes of information, as opposed to hundreds of megabytes common with other search engines (3).

Livelink Spider is the crawler software that locates the documents on

the corporate intranet and external Web sites and locally indexes their full text. Documents from relational databases, flat files, HTML, SGML, and 40 other common office data formats can be indexed. It also has the capacity to index Internet mail files and Internet newsgroups. Livelink Spider can be configured to "crawl" to specific domains, server directories, and file types, and conversely, it can be configured not to crawl to specific domains or server directories. Livelink Search has the flexibility to support multiple indexes on one server and multiple indexes on multiple servers, making decentralized collections of information easily searchable.

Open Text offers a variety of support for their products, including training courses for end-users, administrators, and developers; and online reference information and user guides. Customer Service Representatives are available to support any questions regarding functionality, use, and configuration of Open Text products; however, in order to take advantage of this service, you must subscribe separately to their Customer Assistance Program. At the time of this writing, the price for Livelink Search and Livelink Spider is \$12,000 and \$12,500 per server, respectively. Netscape's Commerce Server communications software is also included in the package.

Full Boolean searching (AND, OR, and NOT) is supported by Livelink Search, as is proximity searching (NEAR), advanced similarity searching (find more results like this one), truncation with a wildcard (*), and full-phrase searching (phrases with no stopwords). Keyword searches look for literal matches of the words, and concept searches use a thesaurus to locate related terms. Searches can also be run to query a specific field of a document (Figure 4). If a Search Application Programmable Interface (API) is purchased and developed for Livelink Search, results can be manipulated for further advanced searching options.

Retrieved results are ranked based on an intelligent ranking algorithm and can be viewed in three different formats: simple ASCII, on-the-fly HTML, or in native format. Livelink Search can convert non-HTML documents on-the-fly so that they can be viewed by any web browser. Document summaries, if not originally provided, can be created using an automatic document summary generator. There is also an option to view the keywords in KWIC mode, where the keywords are highlighted and the user can easily see where the keyword(s) occurs in the retrieved document.

In addition, searches can be restricted to query-specific sections of documents, because the search tool does index documents based on tags or database fields. The client interface is customizable to suit the needs of the users' searching preferences.

Verity SEARCH'97 (<http://www.verity.com>)

Verity, founded in 1988, is the developer of the Topic family of search and retrieval tools for the enterprise and the Internet. In the Fall of 1996, Verity relaunched its entire suite of search and retrieval tools under a new name: SEARCH'97. SEARCH'97 is a comprehensive, flexible platform for deploying search applications across the corporation. The Verity indexing format is being used by over 500 companies worldwide. Some of the companies bundling the search engine into their software include: SAP, Lotus Notes, Individual Inc., Adobe Acrobat, Documentum Inc., Xyvision, Netscape servers, Dow Jones, Reuters, and Ziff Davis

SEARCH'97 can index information from virtually any document format that has been used in the last ten years, including relational databases like Informix, Sybase, and ODI. It is also working towards indexing data from data management files such as Lotus Notes, SAP, Informix, and Documentum. Touted as the mechanism to harness the "corporate memory" of an enterprise, SEARCH'97 facilitates the collection, management, and retrieval of information throughout a corporation and specified sites on the

Internet, and makes the data available at an employee's desktop.

The SEARCH'97 platform includes a variety of components: SEARCH '97 Personal, Information Server, Agent Server, Advanced Search and Query enhancements, and Knowledge application tools and advanced navigation. SEARCH'97 Personal is an interface used to initiate search queries, access search agents, and implement searches. SEARCH'97 Personal can locally index Internet Web sites at the individual's computer so that personal Internet Web sites can also be queried along with remote corporate indexes. It is supported from within a web browser or Microsoft Exchange. Results can be viewed in virtually any file type, even if the native application is not available locally. SEARCH'97 Personal is available for UNIX, Mac, Windows 95 and NT.

At the center of the SEARCH'97 framework is the Information Server. The Information Server indexes and manages corporate information--the "corporate memory"--and uses a web browser or SEARCH'97 Personal as an interface. A web spider is also included to add corporate data and/or Internet sites automatically to the main index.

The full text of documents is indexed; the indexes are updated automatically when data is added, changed, or deleted. The indexing tools support access to virtually any document format including common office document formats, HTML, PDF, and ASCII text. Remote indexing is available to store information from different sites throughout a corporate intranet. The Information Server also acts as the integration point for advanced searching components such as the agent server, enhanced query, visualization, and knowledge and navigation tools. According to Verity's Product Brief, the following platforms support Information Server: Solaris, IBM AIX, HP/UX, Windows NT, DEC Win Alpha, and DEC UNIX.

SEARCH'97 Agent Server automates the search and retrieval process for the individual or corporation. A search profile is prepared and the agent notifies the requester when information or data match the search profile. Individuals customize their information profiles with a set of keywords, specific sources (Internet or intranet sites, including databases) which the query will be run against, and the preferred method for notification (via email, web page, or pager). The agents run continuously and instantaneously alert the user when new information has been added to any of the sources specified in the search profile. Hundreds of thousands of agent profiles can be initiated per server. SEARCH'97 Agent Server currently operates on Solaris 2.5 and Windows NT 3.5.1 platforms. General availability for Agent Server is scheduled for first quarter of 1997 with pricing at approximately \$70,000.

A Technical Support site ([http:// www.verity.com/tech-support/index.html](http://www.verity.com/tech-support/index.html)) is available from the Verity home page. This site includes a technical support information sheet providing phone and fax numbers as well as email addresses for Verity offices worldwide. The technical support information sheet also details the procedures for obtaining technical support. Online information is searchable and includes FAQs, selected data from their Help Desk database, and selected technical notes. Verity also offers a number of educational courses (<http://www.verity.com/educ/index.html>) for their products. Courses are taught at Verity Training Centers in Sunnyvale, CA and Fairfax, VA or can be conducted on-site.

The Verity search engine offers both literal text and Boolean searching capabilities. Other searching options are customizable using standard web forms.

For literal text queries, commas placed between key terms will search on any of those keywords (implied OR). Truncation of words occurs automatically; however, a specific word or phrase can be searched simply by

placing quotation marks around the word or phrase. A wildcard can also be used to find variant letters at the beginning of a word or letters. Field searching is available for querying a specific date or author. Proximity operators are also supported; search terms can be specified to show up "near" each other, in the same phrase, sentence, or paragraph. A thesaurus is available to retrieve synonyms for additional search terms.

Natural language queries and query by example (find me more like...) are also supported. The search engine takes the user's query, whether literal or Boolean, and supplements it with "fuzzy logic"--an operator that calculates a "more the better" score to determine relevancy ranking.

Multiple collections (or indexes) can be searched simultaneously by the individual and are selected from a list or drop-down menu. The user can also determine the number of results returned from the search query.

Documents returned are given a score and listed in order of relevance. Results can be previewed using a rich-text translator, or displayed in the "native format" that can range from data in an Oracle database to Lotus Notes documents to Adobe Acrobat PDF files. Native formats can only be viewed when the requested file format is available locally on the user's computer, or when a suitable viewer is used.

Additional add-on components to the basic SEARCH'97 can increase the flexibility of searching and improve the relevance of results. These optional intelligent search components include: Enhanced Query, Visualization, and Knowledge and Navigation Tools. Enhanced Query uses query technologies such as natural language processing (NLP) and query by example (QBE). A user can type in a search in the form of a question and then use NLP to locate information based on the phrases that were entered. In using QBE, a searcher can copy an example of relevant text from a retrieved result and paste that text in the search form. The QBE engine will then reformulate the search and locate information relevant to the text that was submitted.

The Visualization components (clustering and summarization) make it easier for users to identify relevant information. Clustering organizes the retrieved results into groups based on commonality of terms. The Summarization component creates an overview of individual documents based on an algorithm that determines the significance of the sentences that make up the documents. These summaries are more sophisticated than the typical summaries created by just the document title and following few lines of text.

Navigation tools allow the user to move through documents more easily by using hyperlinks from one document to another. To further facilitate knowledge transfer within an organization, a systems administrator can use Verity's Knowledge Tools. These tools allow administrators to create their own knowledge bases specific to their business environment that include, but also extend beyond the typical functionality of dictionaries and thesauri. These navigation tools would provide more precise search results by filtering out and eliminating irrelevant documents.

ZyIndex Webserver (<http://www.zylab.com/>, <http://www.zylab.nl/>)

ZyLab International, Inc. was founded in 1983 with the introduction of PC-based full-text indexing and retrieval software. Today, ZyLab offers complete web-based publishing and indexing solutions in its ZyIndex Webserver and ZyImage Webserver product lines. ZyIndex Webserver, the software package we will be focusing our attention on in this section, provides full-text indexing of document collections in over 30 formats and makes them searchable through the Internet or corporate intranet. ZyImage Webserver, a companion product to ZyIndex Webserver, combines the ZyImage scanning interface for OCR (optical character recognition) of documents in electronic format with the powerful indexing and search and retrieval

engine of ZyIndex Webserver.

All technical specifications and searching functionality described apply to both ZyIndex and ZyImage Webserver. The main difference between the two products is that the ZyImage Webserver offers users the additional benefit of being able to view images of scanned documents with an easy-to-use scanning interface. ZyIndex Webserver sells for \$5,995 complete, while ZyImage Webserver, which includes the ZyImage OCR and ZyIndex software, sells for \$11,200 (price includes annual update service). Both products are licensed to cover a total intranet site. ZyLabs offers a full range of technical support options including an 800 number, electronic mail, Web site, and support contracts.

ZyIndex Webserver supports the most popular HTTP servers, i.e., Microsoft and Netscape. However, ZyIndex Webserver can be used with any existing web server product that is HTTP 1.0 compliant, running on the Windows NT platform. ZyIndex Webserver provides a proprietary API that handles the search and retrieval process and interfaces with the document index created during the configuration process. In addition, ZyIndex Webserver comes with a set of HTML templates designed to function as the search forms used by end-users through their web browsers, and as the default display format for return and viewing of search results. These templates can be customized to meet client needs.

The ZyIndex Webserver allows clients a great deal of flexibility in indexing features and document and index security. ZyIndex can index document collections located anywhere on the corporate network and can support more than 30 native file formats including all major word processing programs (Word, WordPerfect, etc.), group 4 TIFF, popular database formats (dBase 3 and 4, FoxPro, etc), Lotus, Excel, EPS (encapsulated Postscript), and ASCII and HTML files. However, Adobe Acrobat PDF and Microsoft PowerPoint file formats are not supported at this time.

ZyIndex builds an index based on the documents specified and does not use the documents themselves for retrieval. However, as documents are added or changed, the index is automatically updated. Indexes created by ZyIndex can be very large--up to ten gigabytes can be indexed, or the equivalent of 100 gigabytes of documents. (Indexes of ten gigabytes normally represent approximately 100 gigabytes worth of documents.) If you need to restrict access to certain documents or indexes, ZyIndex allows you to define users and passwords that can be assigned to specific documents, groups of documents, or indexes in order to control security.

Because all indexing done by ZyIndex is on the complete text of the document, and indexes can be very large, a strong set of searching features is needed to ensure accuracy and relevancy of retrieval. ZyIndex Webserver supports Boolean operators and full nesting, phrase searching, advanced proximity searching and truncation, "fuzzy" searches that retrieve words similar to those specified, a "vocabulary" or browse index feature, field searching, and thesaurus for location of synonyms. Searching for numbers or number ranges is supported using standard math operators such as (is less than), (is greater than), =, etc.

In addition, the Concept feature allows web site managers to define searches that cover a particular subject contained in the index, name and save the search strategy, and then display the stored Concept searches for use by end-users searching the index. All of these features are included in an easy-to-use HTML template included with the package.

Retrieved documents, ranked according to relevancy, are automatically translated to HTML on-the-fly for viewing through web browsers regardless of native format, and can then be viewed in native format by launching the appropriate application program. Search terms are highlighted within the context of retrieved documents and users can move "hit to hit" to each

occurrence of the term(s) specified in their search request. Another nice feature of the ZyImage Webserver is the ability to view TIFF images directly through the web browser without the use of a helper application or plug-in, by using a TIFF to GIF converter included with the product.

From the ZyLab home page, you can view a demo of ZyIndex Webserver in action on a test database provided by the National Library of Medicine, as well as use a test database set up by ZyLabs to demonstrate a basic installation using the default searching interface and features.

CHOOSING THE RIGHT INTRANET SEARCH ENGINE FOR THE JOB

As demonstrated by the products reviewed in this article, there are many different things that need to be taken into consideration when evaluating and selecting a search engine for an intranet. Size of the site, the type of documents included, the number of web servers, server platform, and technical expertise available are all major factors influencing the selection of an intranet search engine.

If the intranet site is small and does not contain documents in formats other than HTML and ASCII text, the freeware search engines may be enough to do the job. The frequent downside to these free tools, however, is that advanced technical knowledge is needed to configure and customize the software for site-specific use, and that advanced searching functionality found in the commercial engines is not available. In addition, little formal technical support is offered by any of the free intranet search engines, except for Excite, which charges for its support and maintenance contract.

For large, highly-developed intranet sites, spending the money on a investment. Having the ability to index documents in a variety of file types, including distributed relational databases, and from a variety of locations, both internal and external, makes integrating and then retrieving information on an intranet much easier. Advanced searching features such as field, proximity, and concept searching, as well as the intelligent alerting capabilities promised with the next release of Verity's SEARCH'97 Agent Server, can reduce the number of irrelevant hits produced by a search of a large document collection and automate the search process so that users are automatically notified when content that matches their search profile is added.

WHAT WILL THE FUTURE HOLD?

In this fast-paced, ever-changing world of web-based information retrieval, there are several trends that promise to have a large effect on search and retrieval functionality of intranets.

Not Just Documents

Integration of access to distributed databases (not just documents) with intranet search engines is of paramount importance if intranets are to evolve to the next level of importance in the enterprise.

There are a host of vendors that provide gateways and development tools that can make access to distributed databases from the World Wide Web a reality. Intranet search engine vendors such as Fulcrum, Verity, and Open Text are poised to move to that next level, and may emerge as the favorites for intranet search engines in the near future.

Bundling With Server Software

More and more, web server software packages designed for intranet use are coming bundled with search engines designed to work with the web server software. The two major commercial web server vendors, Netscape and Microsoft, have already capitalized on this trend by including search engines as part of their web server offerings.

Netscape's Enterprise Server comes with the option of purchasing Verity's search engine and using Netscape's Catalog Server (based on Harvest) for indexing document collections. Microsoft's Internet

Information Server provides searching functionality through the Microsoft Index Server, a free package that can index and search HTML and file formats created by the software packages in the Microsoft Office Suite. Although both Netscape's Enterprise Server and Microsoft's Internet Information Server provide a built-in searching solution, other search engine products can be used with these web servers, if desired. As intranets grow, it is likely that even though a basic search engine may be included with a web server product, a separate search engine may be purchased as well, depending on the size and complexity of the intranet site.

Intelligent Agents on Alert

The addition of intelligent agents that can "remember" a search query and run it unattended against both internal and external indexes is another emerging trend that will surely become a favorite with intranet users. Products such as Verity's SEARCH'97 Agent Server and other products mentioned in the sidebar automate the search process and provide vital alerting services that can keep users up-to-date on topics in their areas of interest in "real-time" fashion. The personal search agent products also have the potential, when used with an intranet search engine, to combine results from the inside intranet world with the outside Internet world, giving users a comprehensive view of very current information on specified topics.

Getting It All

Will we ever really be able or even want to search all internal information and external information using one package? Already, typical end-users are becoming frustrated with the amount of retrieval returned by the popular Internet search engines. Intranets, as they grow, have the potential to inspire that same frustration if proper indexing and search and retrieval tools are not developed and implemented.

Gaining balance between providing relevancy, comprehensiveness, and manageability of information on intranets and the Internet as a whole, through development of a set of end-user tools for retrieving and filtering large sets of information, will provide one of the greatest challenges to information professionals in the coming months and years.

REFERENCES

- (1) Levitt, Lee. "Intranets: Internet Technologies Deployed Behind the Firewall for Corporate Productivity." Prepared for the Internet Society INET '96 Annual Meeting. <<http://www.process.com/intranets/wp2.htm>>
- (2) Balderston, Jim. "Search Engine Vendors Eye Intranets: Intranets Mean That Search Tools Must be Fine-Tuned for Corporate Needs." InfoWorld 18, No. 27 (July 1, 1996): pp. 41.
- (3) Canada NewsWire. November 20, 1996.

Communications to the authors should be addressed to Peggy Zorn, Parke-Davis Pharmaceutical Research, 2800 Plymouth Road, Ann Arbor, MI 48150; 313/996-7202; zornm@aa.wl.com and/or Mary Emanoil, Parke-Davis Pharmaceutical Research, 2800 Plymouth Road, Ann Arbor, MI 48150; 313/996-1814; emanoim@aa.wl.com and/or Lucy Marshall, Edge Information Services, 2642 E. Cholla St., Phoenix, AZ 85028; 602/485-9363; edgeinfo@dancris.com and/or Mary Panek, United Technologies Research Center, 411 Silver Lane, MIS 129-01, East Hartford, CT 06108; 860/610-7478; panekmt@utcc.utc.com.

RELATED ARTICLE: The Next Level

SEARCHING DATABASES THROUGH AN INTRANET

While intranet search engines can index and search collections of documents on an intranet, what about including existing databases that a company might have, such as Oracle, Sybase, or Microsoft SQL Server databases? Often, the bulk of the most important information a company has

is stored in one of these formats. Gaining access to this vital information from a corporate intranet is a hot issue and will likely form the next wave of major intranet expansion.

Most intranet search engines are not yet able to integrate the searching of information to this next level: tapping both documents and collections of information stored in traditional database formats. Almost all major database vendors have created web-based interfaces and gateways to their products. There are many generic products available that allow you to interface to ODBC (open database connectivity)-compliant databases that use one development environment.

Below is a listing of major vendors competing in the expanding market of web-based connectivity to ODBC-compliant databases, with URLs for more information. Some, such as Oracle WebSystem and Sybase's NetImpact Dynamo, are targeted to specific database products. However, despite these specializations, all products still claim to have the ability to interface with any ODBC-compliant database product.

ColdFusion <http://www.allaire.com>
Everyware's Tango Enterprise <http://www.everyware.com/>
MEGASOFT Web Transporter <http://www.megasoft.com/>
Microsoft Internet Information Server with Microsoft dbWeb
<http://www.microsoft.com>
Netscape LiveWire <http://www.netscape.com>
NeXT WebObjects <http://www.next.com>
Oracle WebSystem <http://www.oracle.com.sg>
O'Reilly's WebSite Professional <http://www.ora.com/>
Sybase NetImpact Dynamo and web.sql <http://www.sybase.com>
WebDynamics Spider <http://www.w3spider.com>

RELATED ARTICLE: Intelligent Search Agents

Intelligent search agents allow users to create profiles based on their information needs and to simultaneously search selected sites from the external Web, corporate intranet, newsgroups, etc. for the desired information. It is similar to the use of alerting services or SDIs in traditional online searching, except that the intelligent agent can learn from the results, thereby refining the query and returning more valuable information with each new search.

The degree to which intelligent agents are being used varies among software products. Some are simply monitoring tools to alert users when changes have been made to bookmarked sites, but others make associations between search terms and other frequently occurring terms found in search results and then alert the user to these associations. Regardless of the level of agent sophistication, one can expect that software developers will continue to incorporate and improve upon this technology in their products.

Search software that currently uses agent technology includes CyberSearch and WebCompass. Frontier Technologies has announced the release of the 3.0 version of CyberSearch

(<http://www.frontiertech.com/products/cyberseb/csspecl.htm>), its Internet searching and bookmarking utility. Frontier calls the new version of CyberSearch "a global information management tool" because it searches documents on the Internet, intranet, and local PC. Through the use of standard Internet search engines such as Alta Vista, Lycos, Excite, and InfoSeek and server-side indexing of internal documents, this product incorporates the concept of seamless searching among all the information sources accessible to a user.

Quarterdeck Corporation intends to develop a version of its well-reviewed WebCompass software (<http://www.quarterdeck.com/qdeck/products/webcompass/>) that will not only allow users to query multiple search engines, as is the case with its

current release, but will also allow for the inclusion of intranet resources. The current version of WebCompass searches multiple Internet search engines simultaneously, sorts the results, and removes duplicate hits. Results are returned in a Microsoft Access database for easy manipulation.

Other software tools that incorporate intelligent agents and that may be beneficial for multisite searching are available. To keep abreast of new developments in the use of intelligent agents for intranet/Internet searching, visit the Complete Intranet Resource (<http://www.intrack.com/intranet/>). This site provides detailed information about intranets, including a list of software sources.

COPYRIGHT 1997 Online Inc.

SPECIAL FEATURES: table; illustration

DESCRIPTORS: Intranets--Usage; Online searching--Usage

PRODUCT/INDUSTRY NAMES: 7399200 (Info Services ex Database)

SIC CODES: 7389 Business services, not elsewhere classified

FILE SEGMENT: TI File 148